

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université 20 Août 1955 - Skikda

Faculté des Sciences - Département d'Informatique



Mémoire de fin d'études pour l'obtention du diplôme de
Master en informatique.

Option : Génie Logiciel Avancées et Applications (G.L.A.A)

Thème

**EXTRACTIOS DES CONNAISSANCES A
PARTIR D'UN FICHER LOG POUR
ORIENTER DES PARCOURS SUR
APPRINTISSAGE E-LEARNING**

Encadré par :

Dr : BENOUDINA Lazhar.

Réalisé par :

- BERRAHAL A/Malek

Annéeé Universitaire 2021-2022

Sommaire :

| | |
|---|--|
| Introduction générale | |
| Chapitre 01 : Les parcours d'apprentissage dans les environnements d'E-learning..... | |
| 1. Introduction | |
| 2. Les environnements d'apprentissage d'E-learning..... | |
| 2.1. Définition | |
| 2.2. C'est quoi l'apprentissage ?..... | |
| 2.2.1. Définition | |
| 2.3. L'apprentissage ou la formation à distance | |
| 2.4. Les modalités des formations à distance | |
| 2.4.1. Les cours par correspondance | |
| 2.4.1.1. Définition | |
| 2.4.2. Les formations en ligne | |
| 2.5. Le E-learning comme un apprentissage à distance et en ligne | |
| 2.5.1. Définition | |
| 2.5.2. Les principaux avantages de E-Learning | |
| 2.5.3. Les limites du E-learning | |
| 2.5.4. Les acteurs dans E-Learning | |
| 2.5.5. Les éléments (Composants) d'un environnement d'E-learning | |
| 2.5.6. Les types d'interaction en E-Learning | |
| 2.5.7. Les modes de diffusion du E-Learning | |
| 2.6. Les systèmes de gestion d'E-learning | |
| 2.7. Les plates-formes d'E-learning | |
| 2.7.1. Définition | |
| 2.7.2. Les fonctionnalités de la plate-forme | |
| 2.7.3. Les types des plates-formes d'apprentissage | |
| 2.7.4. Les services offerts par les plates-formes d'E-learning | |
| 3. Les parcours d'apprentissage | |
| 3.1. Définition | |
| 3.2. Modélisation | |
| 4. Extraction des parcours d'apprentissage à partir de l'historique d'apprentissage | |
| 4.1. Définition de fichier LOG | |
| 4.2. Définition : Traces..... | |
| 4.3. Prétraitements des traces | |
| 5. Conclusion..... | |

Chapitre 02 : La prédiction des parcours d'apprentissage basé sur les fichiers logs.....

| | |
|--|--|
| 1. Introduction | |
| 2. La prédiction | |
| 2.1. Définition | |
| 2.2. Les modèles de la prédiction | |
| 2.2.1. Modèles prédictifs | |
| 2.2.2. Modèles descriptifs | |
| 2.2.3. Modèles de décision | |
| 3. Analyse prédictive | |
| 3.1. Définition | |
| 3.2. Les étapes de l'analyse prédictive | |
| 4. Les méthodes de la prédiction | |
| 5. Quelques exemples de méthodes prédictives..... | |
| 5.1. Arbre de décision | |
| 5.1.1. Définition | |
| 5.1.2. Exemple d'un arbre de décision | |
| 5.1.3. Algorithme | |
| 5.1.4. Les avantages et les inconvénients de la méthode | |
| 5.2. Les réseaux de neurones | |
| 5.2.1. Définition | |
| 5.2.2. Mise en œuvre | |
| 5.2.3. Les avantages et les inconvénients de la méthode | |
| 5.3. La régression | |
| 5.3.1. Définition | |
| 5.3.2. Principaux modèles de régression | |
| 5.4. Les classifieurs (SVM) | |
| 5.4.1. Définition | |
| 5.4.2. Analyse discriminante | |
| 5.4.3. Les idées clés de classifieurs (SVM) | |
| 5.4.4. Le principe de classifieurs (SVM) | |
| 6. La prédiction basée sur le raisonnement à base de cas | |
| 6.1. Le raisonnement à base de cas | |
| 6.2. Les étapes du processus du raisonnement à base de cas | |
| 6.3. Avantages | |
| 7. Conclusion..... | |

Chapitre 3 : Conception et architecture

| | |
|---|--|
| 1. Introduction | |
| 2. Problématique et Objectif..... | |
| 3. Exigences Fonctionnelles | |
| 4. Le contexte global du système Trails_predictor | |
| 5. L'Architecture du Système Trails_predictor..... | |
| 5.1. Fichier LOG | |
| 5.2. Le module de Traitement log | |
| 5.2.1. Nettoyage | |
| 5.2.2. Filtrage | |
| 5.3. Le module extraction des parcours d'apprentissage | |
| 5.4. Le module de génération des modèles des parcours | |
| 5.5. Le module de prédiction des parcours d'apprentissage | |
| 5.6. Le module d'évaluation des parcours prédits | |
| 6. Conclusion..... | |

Chapitre 4 : Implémentation

| | |
|--|--|
| 1. Introduction | |
| 2. Description des plates-formes utilisée | |
| 2.1. Plate-forme matérielle | |
| 2.2. Plate-forme logicielle..... | |
| 2.2.1. Système d'exploitation..... | |
| 2.2.1.1. Windows 10 | |
| 2.2.2. Outils de développement | |
| 2.2.2.1. Environnement de développement NetBeans | |
| 2.2.3 Langages utilisés..... | |
| 2.2.3.1. Le langage JAVA | |
| 3. Présentation de l'application | |
| 3.2. Page d'accueil | |
| 3.3. S'authentifier | |
| 3.4. Ouverture du fichier log | |
| 3.5. Traitement du fichier log | |
| 3.6. Générer les modèles | |
| 3.7. Le prétraitement du fichier log | |
| 3.8. La prédiction des parcours d'apprentissage | |
| 3.9. Evaluation de la réussite | |
| 4. conclusion | |

Conclusion générale.....

Références.....

Liste des figures et tableaux.....

Introduction Générale :

L'internet et l'application des nouvelles technologies de l'information et de la communication au domaine de l'enseignement ont conduit à l'émergence d'un nouveau mode d'apprentissage appelé : E-learning ou l'apprentissage électronique. Surtout après la pandémie COVID-19 ce mode d'apprentissage s'est accompagné par la disponibilité de plusieurs plateformes assurant un apprentissage en ligne et à distance.

Cependant, ces systèmes de E-learning rendent difficile l'assistance et l'encadrement des apprenants à cause d'un manque de contact directe entre les différents intervenants dans le système d'apprentissage. En effet, les formateurs, les enseignants ou les pédagogues ne se trouvent pas en face à face avec leurs apprenants pour déterminer, à partir de leurs expressions de visage, de leurs questions et interactions, s'ils ont bien assimilé les connaissances présentées. En fait, ils ne peuvent pas bien encadrer les apprenants et suivre leur parcours d'apprentissage réel et voir s'il est conforme ou non au scénario pédagogique préétabli au préalable. Pour cela plusieurs travaux se sont focalisés sur l'étude des comportements des apprenants en se basant sur leur historique d'apprentissage. L'étude de ces comportements d'apprentissage est basée sur une analyse des parcours des apprenants qui sont extraits à partir des fichiers logs générés automatiquement par les plateformes d'apprentissage d'E-learning.

Par ailleurs, et pour garantir un apprentissage réussi pour tous ses apprenants en cours d'apprentissage, l'enseignant à besoin de vérifier si les parcours entrepris par ses apprenants mèneront à une réussite ou à un échec d'apprentissage et cela en vue d'intervenir en juste temps pour assister et supporter les apprenants en difficultés d'apprentissage en leur proposant des conseils, des aides et des recommandations.

Actuellement, le seul moyen qui existe pour valider le succès de l'apprentissage d'un cours est en fin de la formation par un examen final. C'est-à-dire, l'enseignant doit attendre la fin de la formation pour valider la bonne assimilation de son cours par ses apprenants et cela n'est pas très efficace parce que ça sera un peu trop tard pour remédier aux problèmes de l'apprentissage, probablement rencontrés par leurs apprenants lors des différentes sessions d'apprentissage, favorisant ainsi leur désengagement et échec.

Pour rendre les systèmes de E-learning plus performant et garantir un apprentissage réussi, nous proposons de leur intégrer un outil d'aide à la décision dédié à l'analyse et la prédiction des parcours d'apprentissage des apprenants en cours d'apprentissage, à partir de l'historique de

l'apprentissage (fichiers logs). Cet outil permettra aux formateurs que ce soit des enseignants, des tuteurs ou des pédagogues, de suivre l'apprentissage de leurs apprenants et prédire leurs parcours d'apprentissage futurs permettant ainsi l'évaluation précoce de leur comportement d'apprentissage, s'il s'agira d'un comportement positif favorisant la réussite ou négatif amenant à un échec.

Notre travail se focalise plus sur la tâche de prédiction qui complète la tâche d'analyse des parcours d'apprentissage déjà réalisée. En fait, la prédiction des parcours des apprenants en cours d'apprentissage est basée sur les résultats obtenus des analyses des parcours d'apprentissage et plus particulièrement sur les modèles des parcours d'apprentissage découverts à partir de l'historique d'apprentissage (fichiers log) relatif aux apprenants, inscrits au même cours, ayant terminé leur apprentissage.

Le mémoire est organisé comme suit : Une introduction générale, quatre chapitres et une conclusion générale :

- **Chapitre I** : Dans ce chapitre, nous avons abordé les concepts relatifs au domaine de l'apprentissage à savoir, les systèmes d'apprentissage d'E-learning et leurs plateformes, les apprenants et leurs parcours d'apprentissage ainsi nous avons défini qu'est-ce qu'un fichier log ?, trace ?
- **Chapitre II** : Dans ce chapitre nous avons expliqué les différents concepts relatifs aux domaines de la prédiction des parcours d'apprentissage (le modèle prédictif, les méthodes de la prédiction, des exemples des méthodes prédictives) et plus précisément la méthode de raisonnement à base des cas que nous allons appliquer par la suite dans notre contribution.
- **Chapitre III** : Dans ce chapitre, nous avons présenté la démarche suivie pour la prédiction des parcours des apprenants en cours d'apprentissage ainsi que l'architecture et le fonctionnement de notre système dédié à la prédiction des parcours d'apprentissage à partir des fichiers logs générés automatiquement par les systèmes d'E-learning.
- **Chapitre IV** : Ce chapitre consiste à présenter l'implémentation de notre système réalisé. Nous avons aussi décrit brièvement les ressources logicielles et matérielles, les langages de programmation et les outils utilisés pour la réalisation de l'application.

Chapitre 01 : Les parcours d'apprentissage dans les environnements d'E-learning

3. Introduction

L'application des Technologies de l'Information et de la Communication pour l'enseignement (TIC) a donné naissance à une nouvelle forme d'apprentissage appelée «E-Learning ».

Le E-Learning est une approche pédagogique qui fait appel aux technologies de l'information et de la communication et qui permet d'élargir le champ des possibilités et d'aider les apprenants à acquérir les compétences dont ils ont besoin.

Dans ce chapitre, on passera en revue les concepts et les notions relatifs au domaine de E-learning.

4. Les environnements d'apprentissage d'E-learning

2.1. Définition :

Un environnement d'apprentissage est un système qui réalise la synthèse entre, d'une part, les avantages de l'exploration libre et de la construction progressive des objets de connaissance et, d'autre part, l'intérêt du guidage propre aux systèmes tutoriels. L'idée centrale est de permettre à l'apprenant de transformer rapidement et efficacement ses expériences en connaissances organisées [1].

Un environnement d'apprentissage est aussi un système qui privilégie l'idée que la meilleure façon d'apprendre c'est de se retrouver dans une situation réelle de conception et de travail. Plutôt que de construire des logiciels orientés sur l'explicitation formelle de connaissances scolaires, nous pensons qu'il est maintenant possible de concevoir des outils et des environnements qui assisteraient l'apprenant efficacement dans les problèmes auxquels il doit faire face dans sa carrière d'apprenant (Brown, 1989). Il apparaît que l'apprentissage incident de l'ensemble des connaissances nécessaires à la résolution d'une tâche pose moins de problème de motivation ou d'attention, si l'intérêt pour la tâche est assuré à un niveau élevé [1].

Un environnement d'apprentissage est enfin un logiciel qui permet d'entraîner un certain nombre de procédures de "calcul" pour acquérir les automatismes nécessaires à la résolution de problèmes récurrents .Il doit aussi prendre en compte les champs conceptuels du domaine et le

définitions de termes pour donner à l'apprenant la possibilité de communiquer efficacement avec les spécialistes d'un champ de compétence [1].

2.2. C'est quoi l'apprentissage ?

2.2.1. Définition :

L'apprentissage est l'action de transmettre des connaissances nouvelles ou savoirs à un apprenant. Il s'agit du système et de la méthode d'enseigner, composée par tout un ensemble de connaissances, de principes et d'idées transmis à quelqu'un [2].

L'Apprentissage implique l'interaction de trois éléments : enseignant ; apprenant ; et l'objet de connaissance. Tel que, l'enseignant est la source du savoir et l'élève est tout simplement le récepteur illimité de celui-là [3].

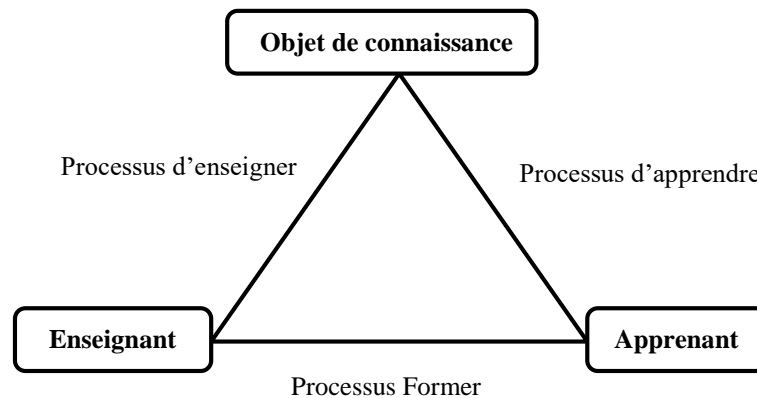


Figure 01 : Le triangle pédagogique de J.Houssage [3]

2.3. L'apprentissage ou la formation à distance :

La formation à distance est un enseignement qui ne comporte pas la présence directe des éléments d'apprentissage sur place mais s'appuie principalement sur un ensemble de supports pédagogiques multimédia [4].

L'apprentissage ou Formation à distance (FAD) représentent de bonnes solutions pour suivre des cours par correspondance à votre rythme, tout en obtenant un diplôme reconnu par l'Etat, ou pour vous améliorer dans une matière... Qu'elles soient techniques, pratiques ou théoriques, ces formations peuvent vous être utiles à tout moment de votre cursus, et vous permettent de suivre des cours en accord avec votre emploi du temps que vous maîtrisez à 100% [5]. Souvent appelées "cours par correspondance", mais aussi "E-learning" ou encore "Mooc" (massive open online course), ces formations effectuées à distance offrent de nombreuses

possibilités de formations, diplômes, soutiens scolaire, préparation à des examens... De plus, la plupart du temps vous pouvez bénéficier d'un stage au sein d'une entreprise qui peut vous aider à l'insertion d'emploi [5].

2.4. Les modalités des formations à distance :

2.4.1. Les cours par correspondance :

2.4.1.1. Définition :

Les cours par correspondance désignent une forme d'instruction, caractérisée par un enseignement dispensé à distance par un établissement et ses professeurs, via des cours, à l'apprenant [6].

Dans le cadre de cours par correspondance, la personne doit disposer d'une adresse postale pour recevoir les documents de formation (DVD, livres, logiciels) afin de retourner ses travaux pour correction en utilisant un service de transport du courrier. La formation est alors effectuée en différé, de manière asynchrone.

Avec l'évolution des technologies, les cours par correspondance sont remplacés peu à peu par les cours en ligne disponibles sur le web. En effet plutôt que de recevoir des cours papier à son domicile, nos formations en ligne attractives et professionnelles sont accessibles n'importe où et n'importe quand. Des évaluations en direct, un suivi personnalisé rendent l'apprentissage encore plus performant [7].

2.4.2. Les formations en ligne :

La [formation en ligne](#) permet aux enseignants d'envoyer des documents aux étudiants de manière simultanée, peu importe la distance. L'apprenant peut télécharger ou consulter en ligne des supports divers. Le cours peut offrir une plus grande [interactivité](#), ce moyen permettant aux étudiants d'avoir des échanges en direct pour poser des questions ou y répondre. L'apprentissage à distance est donc un nouveau moyen pour augmenter la participation et la compréhension des étudiants.

Les principaux intérêts aux développements des formations en ligne sont : le coût, l'absence de contrainte géographique, les disponibilités horaires, une meilleure communication entre l'enseignant et l'étudiant, un suivi.

Grâce à l'informatique, il est possible de pouvoir suivre un cours en temps réel avec l'enseignant et d'acquérir des compétences [7].

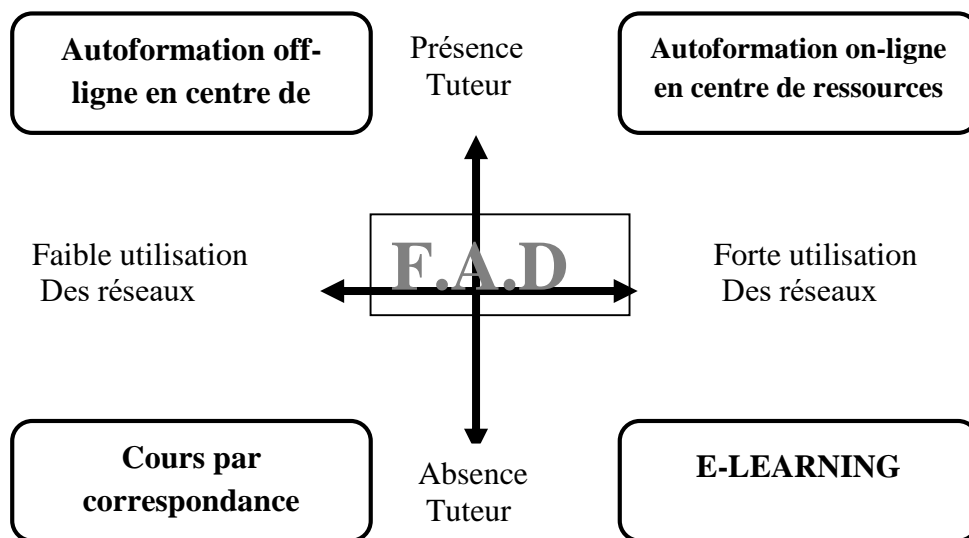


Figure 02 : Une topologie de la formation à distance

2.5. Le E-learning comme un apprentissage à distance et en ligne :

Pour diffuser et gérer leurs processus d'apprentissage, un certain nombre d'organisations et d'institutions d'enseignement utilisent des systèmes d'apprentissage à distance et en ligne dites de E-learning. Mais qu'est-ce que le E-learning ?

2.5.1. Définition :

Plusieurs définitions ont été mises au point, par différents auteurs, pour définir le E-learning. Nous allons citer celles qui sont les plus répandus :

« E-learning, le « E » est une référence explicite aux technologies de l'information, consiste à utiliser les ressources de l'informatique et de l'Internet pour acquérir, à distance, des connaissances. Il peut être défini comme étant une formation effectuée par des moyens de télécommunications en temps réel (synchrone) ou en différé (asynchrone) » [8].

« E-learning est un mode d'apprentissage basé sur l'utilisation des nouvelles technologies, permettant l'accès à des formations en ligne, interactives et parfois personnalisées, diffusées par l'intermédiaire d'Internet, d'un intranet ou autre média électronique, afin de développer les compétences, tout en rendant le processus d'apprentissage indépendant de l'heure et de l'endroit. »[9].

« E-learning est un domaine révolutionnaire, un changement dans le champ de l'apprentissage. Les instructions qu'il offre en ligne peuvent être fournies n'importe quand et n'importe où par une gamme très vaste de solutions d'apprentissages électroniques telles que les groupes de discussions, les cours virtuels « en live », vidéo et audio, Web chat, simulations » [10].

De cela, nous pouvons conclure que le E-learning est un processus d'apprentissage à distance, s'appuyant sur des ressources multimédia qui permet à une ou plusieurs personnes de se former, à leur rythme en fonction de leurs besoins et de leurs disponibilités, à partir de leur ordinateur.

4.5.2. Les principaux avantages de E-Learning :

Le E-learning offre beaucoup d'avantages pour l'apprentissage, nous citerons les plus répondus à savoir [11] :

- **Permet la formation «juste à temps» :**

La formation virtuelle permet aux apprenants de se former et de se rafraîchir la mémoire en tout temps sur une foule de connaissances sans avoir à attendre la diffusion d'un cours à un moment bien précis.

- **Offre une formation en toute souplesse :**

Les modes de diffusion de l'apprentissage virtuel offrent une de large possibilités pour recevoir la formation: dans une salle de classe traditionnelle aménagée en conséquence, dans un bureau réservé spécifiquement à l'apprentissage individuel ou à des stations de travail aménagées un peu partout dans l'entreprise pour les employés n'ayant pas accès à un ordinateur, et même à la maison.

- **Contribue à réduire les coûts de formation :**

L'apprentissage virtuel requiert un investissement initial important, mais son utilisation peut souvent engendrer des économies substantielles. Elles se réalisent principalement par la réduction des frais de déplacement des apprenants, des honoraires des formateurs, des pertes de temps de travail grâce à la diminution du temps requis pour l'apprentissage, des pertes de productivité en rendant accessible la formation sur les lieux de travail.

- **Uniformise et personnalise l'apprentissage :**

L'apprentissage virtuel assure à la fois l'uniformité de la formation et rend accessible aux employés des cours et des parties de cours selon les besoins exprimés par chacun.

- **Permet de diffuser l'apprentissage de manière simultanée à un large auditoire :**

Que vous lanciez sur le marché un nouveau produit ou un nouveau service et que vous deviez former une multitude de personnes simultanément, à un moment précis, l'apprentissage virtuel est la stratégie toute désignée. En effet, vous pouvez joindre, au moment où vous choisissez, tous ceux qui ont accès à un ordinateur branché sur Internet ou sur votre intranet.

Vous pouvez dès lors vous assurer que tous auront la même formation au même moment donné [12].

4.5.3. Les limites du E-learning :

Mais des limites subsistent au cyber apprentissage parmi lesquelles, nous citons [13] :

- **L'absence de contacts humains :**

Et oui, le point fort du E-learning, le virtuel, est aussi son principal point faible : il empêche tout contact humain en face à face. Pour ceux qui travaillent en open space, cela ne pose a priori pas de problème, mais pour les personnes qui sont déjà isolées (télé-travail, petites équipes...), cela peut apparaître comme un véritable frein, ou une difficulté.

- **Confusion multimédia :**

Face à la profusion des modules, certains participants peuvent prendre peur, ou ne pas oser les utiliser, ou ne pas savoir les utiliser, et donc passer plus de temps sur cet aspect technique que sur le cœur de la formation : ce qui doit être transmis ou appris.

- **Le contenu :**

Une fois mis en ligne, le contenu est difficilement adaptable selon le type de public. Il est en effet généralement créé à l'avance. Or tout formateur intervenant sait qu'il devra adapter le vocabulaire ou approfondir telle ou telle partie, en fonction de son public. Dans l'E-learning, c'est figé.

4.5.4. Les acteurs dans E-Learning :

Les acteurs principaux dans un environnement E-Learning sont : apprenants, tuteurs, concepteurs.

Apprenant : désignation de la personne qui se forme ou s'auto-forme dans un dispositif de formation continue. Par opposition à « élève » ou « stagiaire », le terme d'apprenant renvoie à une logique d'implication de la personne dans le déroulement de son parcours de formation individualisée. Le choix de ce terme souligne aussi la prise en considération des caractéristiques personnelles, aux plans conatif, social et professionnel de l'apprenant, par les formateurs/tuteurs en charge de la mise en œuvre et de l'accompagnement de la formation.

Tuteur : (coach) un tuteur est un agent responsable de transmettre des connaissances à un apprenant dans le cadre d'une fonction appelée tutorat (assistance et soutien dispensés aux étudiants par des enseignants). Dans le cadre d'une formation à distance, le tuteur occupe plusieurs fonctions :

- il détermine avec l'apprenant les objectifs de la formation,
- il constitue les groupes,
- il assure le suivi pédagogique de la formation (réponses aux questions des apprenants, analyse de la progression, conseils personnalisés).

Le tuteur joue un rôle moteur dans la formation. La qualité du suivi permet de garantir la motivation de l'apprenant et d'éviter qu'il abandonne sa formation en cours de route...

Enseignant : c'est le fournisseur de contenus pédagogiques interactifs (cours).

Administrateur : désignation de la personne qui contrôle, dirige et commande les différents dispositifs dans la plateforme d'E-learning. L'administrateur de la plate-forme peut s'occuper de :

- la création des comptes utilisateurs,
- la mise en ligne des cours et affectation des professeurs,
- la modération des forums,
- l'analyse des logs contenant l'intégralité des actions effectuées au sein de la plate-forme.

4.5.5. Les éléments (Composants) d'un environnement d'E-learning :

Un système de E-learning est composé de [9] :

- Une communauté d'apprenants
- Une plate-forme d'apprentissage
- Des tuteurs ou animateurs
- Des contenus textuels ou multimédia didactiques
- Une stratégie pédagogique et tutorale
- Des activités de validation de connaissances

Le schéma ci-dessous synthétise l'ensemble des éléments d'un dispositif de formation à distance [14] :

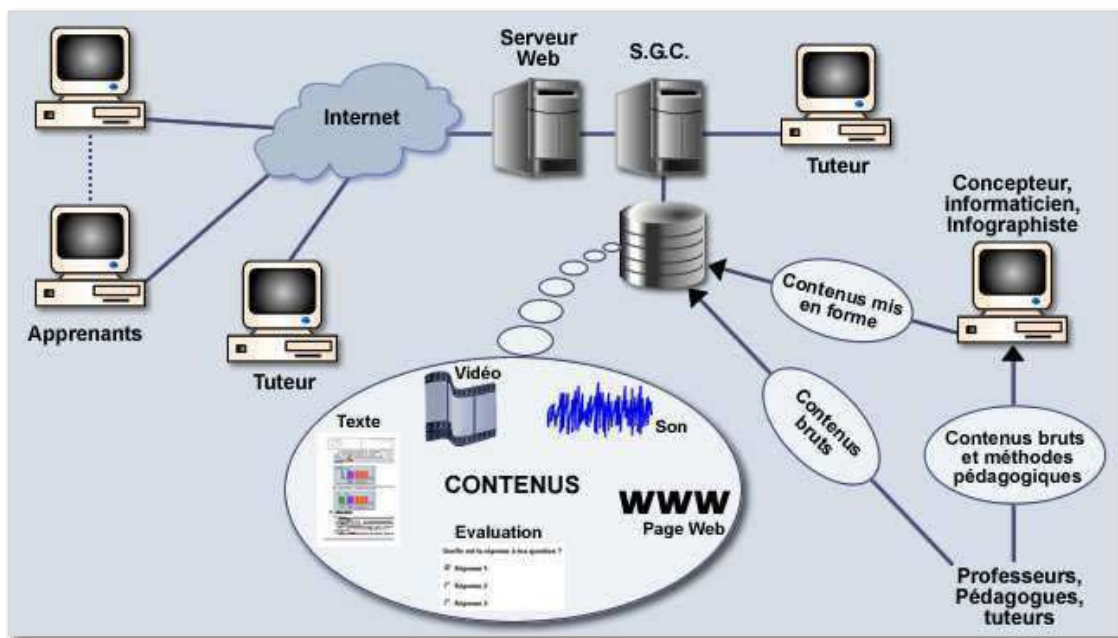


Figure 03 : Les éléments de E-Learning [14]

4.5.6. Les types d'interaction en E-Learning :

L'interaction en E-Learning peut avoir lieu entre [15] :

- **Enseignant-apprenant** : le rôle de l'enseignant dans ce type d'interaction est, soit de transmettre directement les connaissances à l'apprenant, soit le guider pour arriver à la connaissance.
- **Apprenant-apprenant** : Ce type d'interaction permet d'échanger les informations entre les apprenants, de se collaborer entre eux pour réaliser des projets en communs, de discuter et d'échanger des messages...

- **Apprenant-contenu** : L'apprenant peut lire les documents de la formation définis par le système d'E-learning ou définis par son enseignant, comme il peut chercher et lire des contenus publiés sur le Web.

4.5.7. Les modes de diffusion du E-Learning :

Suivre une formation en E-Learning ne signifie pas être seul face à un ordinateur, sans personne pour échanger sur les concepts abordés au cours de la formation ou pour vous apporter un support technique ou pédagogique.

En effet, il existe de nombreuses possibilités de communication en E-Learning, différentes de celles que l'on utilise en formation en présentiel (Chat, forum....)

De façon simplifiée, on distingue 3 modes de diffusion du E-Learning [11] :

- + **Asynchrone** : c'est une méthode de formation en temps différé qui permet à l'apprenant d'accéder à la formation à sa guise et autant de fois qu'il le désire (contenus sous forme de textes, animations multimédias...). Le dialogue entre les formateurs et les apprenants peut se faire via e-mail, messagerie instantanée et forums de discussion, SMS, Partage de documents.

- + **Synchrone** : c'est une méthode en temps réel basée sur la parole et l'écoute.

L'apprenant est en contact avec son formateur et les autres membres de sa classe virtuelle par partage d'application ou visioconférence. Les étudiants connectés posent leurs questions et reçoivent la réponse en temps réel, quel que soit la distance à condition bien sûr que la bande passante soit à la hauteur. Le dialogue entre les formateurs et les apprenants peut se faire via : Chat, Webcam, Micro, Partage d'applications, Fonctionnalités de prise en main à distance.

- + **Mixte** : c'est une méthode qui combine les éléments asynchrones et synchrones. Il peut aussi être utilisé pour homogénéiser les savoirs avant une présentation, une conférence, une intervention ou une formation.

2.6. Les systèmes de gestion d'E-learning :

Les cours en ligne sont créés, gérés et utilisés grâce à des systèmes de gestion d'E-learning appartenant aux catégories suivantes :

- + **LMS - Learning Management System** :

LMS est une famille de systèmes, qui permet la gestion de toutes les activités de la formation. Les systèmes LMS catégorisent les utilisateurs, leur donnent certaines autorisations à des modules de formation et affectent les utilisateurs à des groupes spécifiques de formation [16].

✚ CMS - Content Management System :

CMS est une famille de systèmes d'E-learning dédiés à la création, le stockage, la gestion et la présentation des contenus. L'important dans ce modèle est de fournir des possibilités de réutilisation des objets d'apprentissage, ou des composants de contenu [16].

✚ LCMS - Learning Content Management System :

LCMS est une famille de systèmes d'E-learning les plus avancés technologiquement. Les systèmes LCMS intègrent des fonctionnalités des LMS et des CMS, assurant à la fois la création et la gestion des contenus éducatifs. Ils offrent la possibilité d'évaluer les connaissances assimilées par les utilisateurs [16].

✚ VCS –Virtual Classroom System :

Les VCS sont des systèmes qui incluent les fonctionnalités suivantes [17] :

- Transmission de la voix et de la vidéo en temps réel entre tous les participants.
- Tableau Blanc partagé (shared whiteboard).
- Espace intégré pour la projection de diapositives ou autres supports visuels.
- Capacité d'interaction textuelle, y compris les conversations ou « note-passing».
- Des moyens permettant aux apprenants d'indiquer qu'ils ont des questions.

✚ ITS- Intelligent Tutoring Systems :

Les ITS sont similaires aux LMS. Ils peuvent donner des feedback intelligents à l'utilisateur. Les ITS emploient les techniques de l'intelligence artificielle, pour pouvoir comprendre, informer et diriger l'apprenant quand il termine ses exercices ou ses tests. Ils visent à répliquer le rôle du tuteur qui guide et forme les apprenants d'une manière efficace. Le tuteur humain est souvent remplacé par des entités intelligentes, appelées « agents » capables de suivre et guider l'apprenant durant son apprentissage [18].

2.7 Les plates-formes d'E-learning :

Dans le E-learning la connexion entre le réel et le virtuel s'effectue à travers des plates-formes.

2.7.1 Définition :

Une plate-forme de E-learning appelée parfois LMS (Learning Management System) est considérée comme étant :

- Un site web qui héberge du contenu didactique et facilite la mise en œuvre de stratégies pédagogiques [19].
- Un outil de diffusion et de gestion des connaissances, associant des contenus de cours à des moyens de communication, des outils d'entraînement et d'évaluation [20].
- Un produit dérivé des logiciels CMS (content management system) mais présente des fonctions importantes pour la pédagogie et l'apprentissage. C'est un logiciel de création et de gestion de contenus pédagogiques destinés à trois types d'utilisateurs : l'enseignant, l'étudiant et l'administrateur. Il regroupe les outils nécessaires aux trois intervenants permettant d'incorporer des ressources pédagogiques multimédias, de participer à des activités et d'effectuer un suivi en mode connecté ou déconnecté suivant les paramètres de la plate-forme.

2.7.2 Les fonctionnalités de la plate-forme :

Les principales fonctionnalités d'une plate-forme sont :

- Création de cours et de tests.
- Gestion de documents pédagogiques (indexation, classification, mises à jour,...).
- Gestion d'un espace de travail coopératif entre apprenants et/ou entre enseignants.
- Suivi de l'apprentissage de l'apprenant et son évaluation.
- Gestion de la scolarité de l'apprenant.
- Mise à la disposition de l'apprenant d'outils de travail (éditeurs spécifiques, Outils de téléchargement, outils de simulation de T.P,...).
- Mise à la disposition des différents acteurs d'outils de communications ainsi que les procédés de leurs gestions (forums, messagerie, chat, visioconférence,...).

2.7.3 Les types des plates-formes d'apprentissage :

Il existe un grand nombre de plates-formes d'enseignement à distance sur le marché international, environ plus de 200 dont une trentaine sous licences libres. Parmi les plates-formes nous citons :

- **Plates-formes sous licence libre (ou licence GPL) :** nous pouvons citer : Moodle, Claroline, Ganesha.

➤ **Plates-formes sous licence propriétaire** : e-doceo, myTeacher, Blackboard (nouveau nom depuis 2006 de WEBCT).

2.7.4. Les services offerts par les plates-formes d'E-learning :

Dans le cadre des technologies web, les plates synchrones ('en temps réel'), asynchrones ('en différé'), collaboratives ou de knowledge management. Parmi les services offerts, nous citons [9] :

- La présentation du catalogue de cours.
- L'intégration des modules de cours et des ressources annexes.
- La diffusion des cours à la demande sur les postes des utilisateurs.
- La gestion des activités des tuteurs.
- La gestion des évaluations des apprenants.
- La gestion des processus qualité.
- La construction des parcours de formation personnalisés.
- Le support à la création d'une communauté d'utilisateurs.
- Le reporting pour les ressources humaines.
- La gestion administrative des apprenants incluant les inscriptions et la comptabilité.

5. Les parcours d'apprentissage

3.1. Définition :

Le parcours d'apprentissage de l'apprenant est considéré comme étant un parcours de navigation sur le web puisque les systèmes d'apprentissage sont considérés comme étant des applications web dédiées à l'apprentissage [48].

Le parcours d'apprentissage représente le chemin de la suite d'actions constituant les activités d'apprentissages réalisées en cours des différentes sessions, et cela selon un ordre temporellement déterminé suivant ou non un scénario d'apprentissage préétabli à priori par l'enseignant appelé parcours pédagogique. En fait, le parcours pédagogique permet de décrire les activités d'apprentissage qui seront proposées à l'apprenant et de définir leur articulation dans le dispositif pédagogique, ainsi que les productions qui sont attendues de la part des apprenants [49].

Dans le système d'apprentissage, un parcours d'apprentissage peut-être identifier à travers l'observation de l'interaction des apprenants avec l'environnement d'apprentissage [48].

6. Extraction des parcours d'apprentissage à partir de l'historique d'apprentissage

Le concept de l'historique des événements dans les systèmes d'apprentissage, désigne l'enregistrement séquentiel dans des fichiers, de tous les événements effectués par tous les intervenants à savoir les apprenants. Le journal (en anglais log file ou plus simplement log), désigne alors le fichier contenant ces enregistrements qui sont datés et classés selon un ordre chronologique. Les enregistrements permettent d'analyser pas à pas l'activité de l'apprenant et ses interactions avec son environnement d'apprentissage.

Dans le système d'apprentissage, un parcours d'apprentissage peut-être identifier à travers l'observation des interactions des apprenants avec l'environnement d'apprentissage [48]. Cette observation de l'apprentissage est assurée par les services de suivi des plates-formes d'E-learning.

Le résultat de ce suivi est enregistré sous formes d'observables dans des fichiers logs puis interprétés en traces.

4.1. Définition de fichier LOG :

Il s'agit d'un fichier comprenant différentes informations liées à l'utilisation d'un serveur, d'une application, d'un logiciel ou d'un système informatique. Un **fichier log** peut contenir certaines données confidentielles sur l'utilisateur [21].

Un log (ou fichier log) se présente sous la forme d'un fichier texte classique, reprenant de façon chronologique, l'ensemble des événements qui ont affecté un système informatique et l'ensemble des actions qui ont résulté de ces événements.

Ainsi, pour un serveur de type Web, le fichier log regroupe à la fois les demandes d'accès à chacun des fichiers du serveur :

- Date et heure précise de la tentative d'accès.
- [Adresse IP](#) du client ayant réalisé cet accès.
- Fichier cible.
- Et éventuellement système d'exploitation et [Navigateur](#) utilisé pour cet accès.

Le fichier contient également la réponse fournie par le serveur à cette demande d'accès (si le fichier est trouvé, le poids de celui-ci... sinon, le type d'erreur rencontré) [22].

Le fichier Log généré par les systèmes d'apprentissage en ligne désigne est un fichier de l'historique de l'apprentissage qui contient une liste des observés qui sont automatiquement collectés lors des différentes interactions des utilisateurs avec le système. Chaque observé est constitué principalement des attributs relatifs à la date et l'heure de la collecte,

l'identifiant de l'utilisateur (nom ou adresse IP) et l'action effectuée sur une entité donnée. Les observés peuvent contenir d'autres attributs.

5.2. Définition : Traces

On appelle trace une collection d'observés temporellement situés. Par le terme observé, nous dénotons toute information structurée issue de l'observation d'une interaction, et par séquence temporelle, nous dénotons l'existence d'une relation d'ordre organisant ces observés par rapport à une extension temporelle qui peut être [23] :

- Soit un intervalle temporel déterminé par deux dates, appelées « date de début » et « fin de l'observation » ;
- Soit une séquence d'éléments quelconques fournissant la relation d'ordre temporelle (par exemple l'ensemble des entiers naturels).

Chaque trace possède sa propre structure et sémantique, et ceci suivant le système qui l'a généré et la nature de l'activité tracée. Donc, pour pouvoir être utilisée, une trace est toujours accompagnée de son modèle. On parle ici de modèle de trace.

4.3. Prétraitements des traces :

La collecte des données fournit des données qualifiées de « traces brutes » ou encore « traces primitives » difficilement exploitables en tant que telles [24]. A titre d'exemple [25], les fichiers

log d'un serveur Apache fournissent des centaines de lignes de type :

```
194.78.232.8 -- [10/Jan/2022:15:33:43 +0200] "Get /orion/liens.htm
HTTP/1.1" 200 1893 "http://www-sop.inria.fr/orion/index.html"
"Mozilla/4.0 (compatible ; MSIE 5.0b1; Mac_PowerPC)
```

On note les informations suivantes :

- 194.78.232.8 est l'adresse IP de l'utilisateur ;
- la requête a été reçue le 10 Janvier 2022 à 15 heures, 33 minutes et 43 secondes ;
- l'URL demandée est /Orion/liens.htm (il s'agit d'une URL à laquelle on doit ajouter le nom du serveur, ici www-sop.inria.fr) ;
- la requête a été traitée sans erreur (c'est le statut 200) ;
- le document renvoyé contenait 1893 octets ;

- le lien vers ce document a été trouvé dans le document d'URL <http://www.sop.inria.fr/orion/index.html> ;
- enfin, l'utilisateur travaillait sur un Macintosh (Mac PowerPC) avec le logiciel Mozilla en version 4.0.

Afin de rendre possible l'exploitation de ces traces, il est nécessaire d'effectuer des prétraitements dans le but de les structurer et de faciliter leur exploitation. Cette phase peut être scindée en trois étapes : fusion, filtrage et structuration [26] :

- **Phase de Fusion** : cette phase est nécessaire pour obtenir une vision globale de l'activité lorsque les sources de données observées sont diverses (audiovisuelles, fichiers logs, observation manuelle).
- **Phase de filtrage** : Les phases de collecte et de fusion produisent généralement une très grande masse de données. Afin de raffiner l'analyse, il est possible de réduire cette quantité de données, en considérant seulement les données répondant aux objectifs de l'observation.
- **Phase de structuration** : la représentation des traces avec des paramètres de bas niveau rend le processus d'interprétation difficile. Il est donc nécessaire de procéder à la structuration et la transformation de ces données. A ce stade, les traces collectées sont regroupées, réarrangées et annotées pour faciliter leur interprétation. L'exécution de cette étape ne peut se faire que s'il est possible de préciser comment comprendre et utiliser les traces brutes. Un modèle de traces est donc nécessaire pour définir ce qui est tracé (définir a priori quelles interactions tracer) et/ou filtrer les entités observées (définir a posteriori quelles informations garder).

6. Conclusion

Dans ce chapitre, nous avons abordé les concepts relatifs au domaine de l'apprentissage à savoir, les systèmes d'apprentissage d'E-learning et leurs plates-formes, les apprenants et leurs parcours d'apprentissage ainsi nous avons défini qu'est-ce qu'un fichier log ?, trace ?

Dans le chapitre suivant, nous allons survoler les différents concepts relatifs aux domaines de la prédiction des parcours d'apprentissage et plus précisément la méthode de raisonnement à base des cas que nous allons appliquer par la suite dans notre contribution.

Chapitre 02 : La prédiction des parcours d'apprentissage basé sur les fichiers logs

3. Introduction

Dans les systèmes d'apprentissage de E-learning, un parcours d'apprentissage sous-entend le chemin parcouru par l'apprenant dans une situation d'apprentissage. La prédiction des parcours des apprenants en cours d'apprentissage permet aux enseignants de prédire la réussite ou l'échec de leur apprentissage. Dans le cas de l'échec probable, l'enseignant peut intervenir dès le début du cours pour remédier aux situations problématiques favorisant cet échec et cela en proposant des conseils, des recommandations et des aides à ses apprenants assurant ainsi un apprentissage réussi et personnalisé.

Dans ce chapitre, Nous allons passer en revue la définition des différents concepts de la prédiction des parcours d'apprentissage dans le système E-learning (c'est quoi une prédiction ?, quelles sont les types de prédiction ?, quelles sont les méthodes de prédiction ? ...).

4. La prédiction

4.1. Définition :

La prédictibilité désigne la possibilité que certains événements ou phénomènes soient prévus ou prédits à partir d'une hypothèse ou d'une théorie et de conditions initiales appropriées [26].

La prédiction est une fonction qui permet d'estimer la demande future des biens et services offerts, cette fonction est établit soit mathématiquement (données historiques), soit intuitivement (connaissance), soit en combinant les deux méthodes [27].

2.2. Les modèles de la prédiction :

3.2.1. Modèles prédictifs :

Un modèle prédictif est un modèle de traitement des données, basé sur des algorithmes et des données dans un but : anticiper les comportements et les actions futures [31].

Dans le domaine du marketing, Les modèles prédictifs analysent les performances passées pour estimer la probabilité que peut avoir un client de montrer un comportement spécifique dans le futur afin d'améliorer l'efficacité marketing. Cette classe de modèles comprend aussi des modèles recherchant des motifs de données subtiles pour répondre aux questions de la performance client, tels que les modèles de détection des fraudes. Les modèles prédictifs effectuent souvent des calculs durant les transactions en ligne, pour par exemple, évaluer le risque d'un client donné ou d'une transaction donnée, dans le but d'aider à la décision [32].

Dans le domaine de l'apprentissage, les modèles prédictifs aideront à prédire les comportements futurs des apprenants en cours d'apprentissage....

3.2.2. Modèles descriptifs :

Les modèles descriptifs quantifient les relations entre les données d'une manière souvent employée pour classer les clients ou les prospects dans des groupes. À l'opposé des modèles prédictifs qui se concentrent sur la prédiction d'un seul comportement de client (tel que le risque client), les modèles descriptifs identifient plusieurs relations entre les clients ou entre produits. Les modèles descriptifs n'ordonnent pas les clients par la probabilité qu'ils ont d'entreprendre telle ou telle action, comme le font les modèles prédictifs. Ils peuvent être utilisés, par exemple, pour catégoriser les clients par leurs préférences et leur étape de développement. Les outils des modèles descriptifs sont employés pour créer d'autres modèles pouvant simuler de grands nombres d'agents individualisés et proposer des prédictions [50].

3.2.3. Modèles de décision :

Les modèles de décision décrivent des relations entre tous les éléments d'une décision - les données connues (y compris les résultats des modèles prédictifs), la décision et les résultats attendus de la décision - afin de prédire les résultats des décisions mettant en jeu un nombre de variables important. Ces modèles peuvent être utilisés dans l'optimisation, la maximisation de certains effets et la minimisation d'autres effets. Les modèles de décision sont généralement utilisés pour développer une logique de décision ou un ensemble de règles métier qui produiront l'action souhaitée pour chaque client ou dans chaque circonstance [51].

4. Analyse prédictive

7.1. Définition :

L'analyse prédictive englobe une variété de techniques issues des statistiques et d'extraction de connaissances à partir de données qui analysent des faits présents et passés pour faire des

hypothèses prédictives sur des événements futurs [28]. Dans le monde des affaires, des modèles prédictifs exploitent des schémas découverts à l'intérieur des ensembles de données historiques et transactionnelles pour identifier les risques et les opportunités. Les Modèles capturent les relations entre de nombreux facteurs permettant l'évaluation des risques ou celle des opportunités associés à un ensemble spécifique de conditions, afin d'orienter la prise de décision vers telle ou telle action.

L'analyse prédictive, considérée comme un type d'exploration de données [29], est un domaine de l'analyse statistique qui extrait l'information à partir des données pour prédire les tendances futures et les motifs de comportement. Le cœur de l'analyse prédictive se fonde sur la capture des relations entre les variables explicatives et les variables expliquées, ou prédites, issues des occurrences passées, et l'exploitation de ces relations pour prédire les résultats futurs [30]. Toutefois, il est important de noter, que l'exactitude et l'utilité des résultats dépendent grandement du niveau de l'analyse des données et de la qualité des hypothèses.

7.2. Les étapes de l'analyse prédictive :

Avant d'établir une quelconque prédiction, il est nécessaire de passer par diverses étapes permettant de construire, dans un premier temps, le modèle prédictif, pour pouvoir ensuite l'utiliser et en tirer des réponses [33].

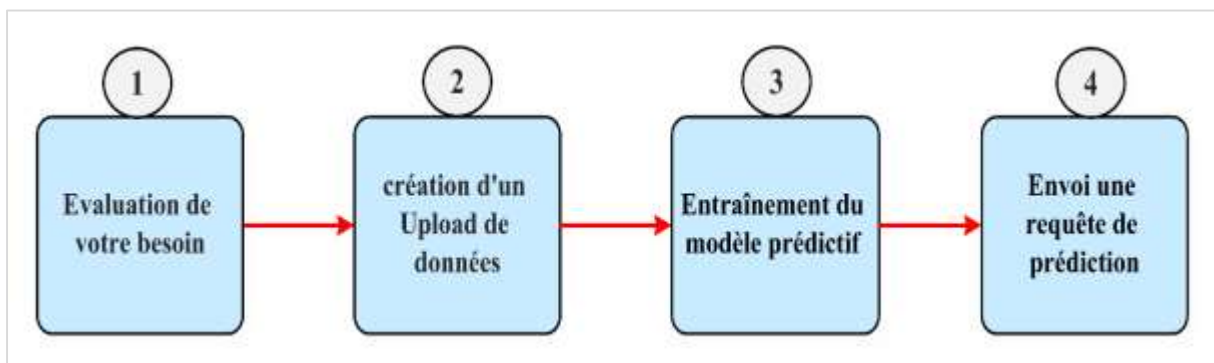


Figure 05 : utilisation de modèle prédictif

- ✚ **Évaluation des besoins :** Vous vous posez une question et vous avez besoin d'y répondre. L'étape de l'évaluation des besoins, permet de valider la pertinence de votre interrogation ainsi que son adaptabilité au modèle prédictif.
- ✚ **Créer l'upload de données :** Sur la base de votre interrogation, les données nécessaires pour y apporter une réponse adaptée, sont sélectionnées. La création de ce modèle de données est l'étape la plus importante et la plus fastidieuse de l'opération. Il est primordial

de bien définir quelles sont les données utiles, quelles sont les variables nécessaires au modèle, et cela prend un certain temps.

✚ **Entraînement du modèle prédictif** : L'upload de données effectué, le système de prédiction va ensuite retourner le modèle afin de l'habituer à recevoir une requête / question.

✚ **Envoyer une requête de prédiction** : Le système de prédiction renvoi une réponse qui est : soit une valeur numérique estimée, soit une catégorisation de l'objet de requête, en fonction des données d'entraînement.

8. Les méthodes de la prédiction

L'objective méthode prédictive est d'expliquer ou de prévoir un ou plusieurs phénomènes observables et effectivement mesurés. Concrètement, elles vont s'intéresser à une ou plusieurs variables définies comme étant les cibles de l'analyse. En exploration de données prédictive, il y a deux types d'opérations [34] :

- ✓ **La discrimination ou classement** : discrimination s'intéresse aux variables qualitatives.
- ✓ **La régression ou prédiction** : s'intéresse aux variables continues.

Les méthodes de classement et de prédiction permettent de séparer des individus en plusieurs classes. Si la classe est connue au préalable et que l'opération de classement consiste à analyser les caractéristiques des individus pour les placer dans une classe, la méthode est dite « supervisée » [35].

Dans le cas contraire, on parle de méthodes « non-supervisées », ce vocabulaire étant issu de l'apprentissage automatique.

9. Quelques exemples de méthodes prédictives

9.1. Arbre de décision :

9.1.1. Définition :

En théorie des graphes, un arbre est un graphe non orienté, acyclique et connexe. L'ensemble des nœuds se divise en trois catégories :

- Nœud racine (l'accès à l'arbre ce fait par ce nœud)
- Nœuds internes : les nœuds qui ont des descendants, qui sont à leur tour des noeuds
- Nœuds terminaux (ou feuilles) : nœuds qui n'ont pas de descendants

Les arbres de décision (AD) sont une catégorie d'arbres utilisée dans l'exploration de données et en informatique décisionnelle. Ils utilisent une représentation hiérarchique de la structure des données sous forme des séquences de décision (tests) en vue de la prédiction d'un résultat ou d'une classe. Chaque individu, qui doit être attribué à une classe, est décrit par un ensemble de variables qui sont testées dans les nœuds de l'arbre. Les tests s'effectuent dans les nœuds internes, et les décisions sont prise dans les nœuds feuille [36].

9.1.2. Exemple d'un arbre de décision :

Voici un exemple ou on doit classifier une situation ou individu en suivant une séquence de tests. Le processus de décision est équivalent à une descente dans l'arbre (de la racine vers une des feuilles) : à chaque étape un attribut est testé et un sous arbre est choisi, ou sinon la parcours s'arrête dans une feuille (une décision est prise) [36].

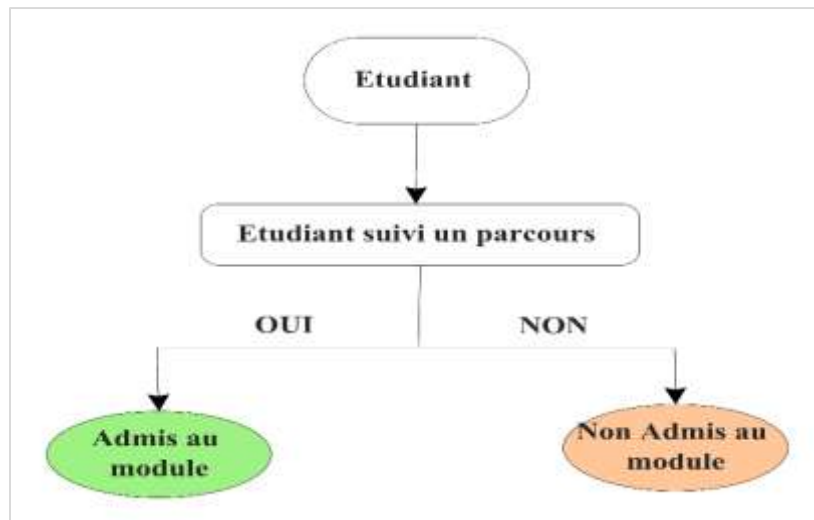


Figure 06 : représentation d'un arbre de décision

9.1.3. Algorithme :

L'algorithme d'apprentissage par l'arbre de décision est décrit comme suit :

Donnée : un échantillon S de m enregistrements classés $(x, c(x))$

Initialisation : $A \leftarrow$ arbre vide

nœud_courant \leftarrow racine

échantillon_courant $\leftarrow S$

Répéter

Décider si le nœud courant est terminal

| |
|---|
| <p>Si (nœud _courant est terminal) alors</p> <p>Étiqueter le noeud courant par une feuille</p> <p>Sinon</p> <p>Sélectionner un test :</p> <p>Créer les fils</p> <p>les échantillons sortants du nœud</p> <p>Fin si</p> <p>nœud _courant ← un nœud non encore étudié de A</p> <p>échantillon _courant : échantillon atteignant nœud _courant</p> <p>Jusqu'à (nœud _courant = \emptyset)</p> <p>élaguer l'arbre de décision A obtenu</p> <p>Sortie : l'arbre A élagué</p> |
|---|

Table 01 : Algorithme d'apprentissage par arbre de décision

9.1.4. Les avantages et les inconvénients de la méthode :

➤ **Avantages :**

Selon [37], les principaux avantages des arbres de décision sont :

- **Adaptabilité aux attributs de valeurs manquantes** : les algorithmes peuvent traiter les valeurs manquantes (descriptions contenant des champs non renseignés) pour l'apprentissage, mais aussi pour la classification.
- **Bonne lisibilité du résultat** : un arbre de décision est facile à interpréter et à la représentation graphique d'un ensemble de règles. Si la taille de l'arbre est importante, il est difficile d'appréhender l'arbre dans sa globalité. Cependant, les outils actuels permettent une navigation aisée dans l'arbre (parcourir une branche, développer un nœud, élaguer une branche) et, le plus important, est certainement de pouvoir expliquer comment est classé un exemple par l'arbre, ce qui peut être fait en montrant le chemin de la racine à la feuille pour l'exemple courant.

- **Traitement de tout type de données** : l'algorithme peut prendre en compte tous les types d'attributs et les valeurs manquantes. Il est robuste au bruit.
- **Sélectionne des variables pertinentes** : l'arbre contient les attributs utiles pour la classification. L'algorithme peut donc être utilisé comme pré-traitement qui permet de sélectionner l'ensemble des variables pertinentes pour ensuite appliquer une autre méthode.
- Donne une classification efficace : l'attribution d'une classe à un exemple à l'aide d'un arbre de décision est un processus très efficace (parcours d'un chemin dans un arbre).
- **Disponibilité des outils** : les algorithmes de génération d'arbres de décision sont disponibles dans tous les environnements de fouille de données.
- **Méthode extensible et modifiable** : la méthode peut être adaptée pour résoudre des tâches d'estimation et de prédiction. Des améliorations des performances des algorithmes de base sont possibles grâce aux techniques qui génèrent un ensemble d'arbres votant pour attribuer la classe.

➤ **Inconvénients :**

Selon [37], il existe quelque limite parmi lesquelles, nous citons :

- **Méthode sensible au nombre de classes** : les performances tendent à se dégrader lorsque le nombre de classes devient trop important.
- **Manque d'évolutivité dans le temps** : l'algorithme n'est pas incrémental, c'est-à-dire, que si les données évoluent avec le temps, il est nécessaire de relancer une phase d'apprentissage sur l'échantillon complet (anciens exemples et nouveaux exemples).

9.2. Les réseaux de neurones :

9.2.1. Définition :

Les réseaux de neurones sont des outils très utilisés pour la classification, l'estimation, la prédiction et la segmentation. Ils sont issus de modèles biologiques, sont constitués d'unités élémentaires (les neurones) organisées selon une architecture [42].

Un nœud reçoit des valeurs en entrée et renvoie 0 à n valeurs en sortie. Toutes ces valeurs sont normalisées pour être comprises entre 0 et 1 (ou parfois entre -1 et 1), selon les bornes de la fonction de transfert. Une fonction de combinaison calcule une première valeur à partir des nœuds connectés en entrée et poids des connexions. Dans les réseaux les plus courants, les perceptrons, il s'agit de la somme pondérée $\sum n_i p_i$ des valeurs des nœuds en entrée. Afin de

déterminer une valeur en sortie, une seconde fonction, appelée fonction de transfert (ou d'activation), est appliquée à cette valeur. Les nœuds de la couche d'entrée sont triviaux, dans la mesure où ils ne combinent rien, et ne font que transmettre la valeur de la variable qui leur correspond [37]. Un nœud de perceptron se présente donc comme suit :

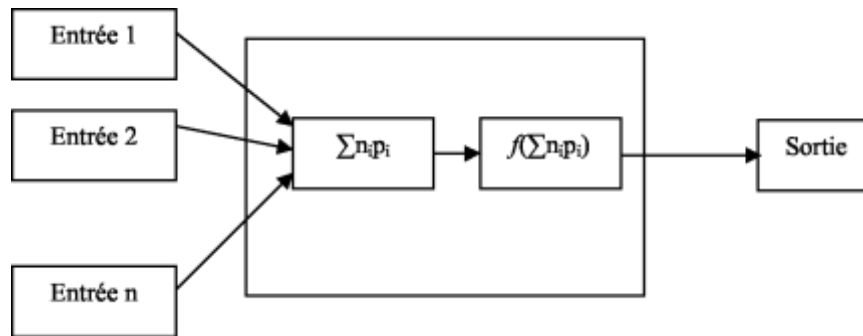


Figure 07 : Nœud d'un réseau de neurone [37].

- n_i est la valeur du nœud i du niveau précédent (la sommation sur i correspond à l'ensemble des nœuds du niveau précédent connectés au nœud observé)
- p_i est le poids associé à la connexion entre i et le nœud observé.
- f est la fonction de transfert associée au nœud observé.

9.2.2. Mise en œuvre :

De façon générale, les étapes dans la mise en œuvre d'un réseau de neurones pour la prédiction ou le classement sont [37] :

1. l'identification des données en entrée et en sortie,
2. la normalisation de ces données,
3. la constitution d'un réseau avec une structure adaptée,
4. l'apprentissage du réseau,
5. le test du réseau,
6. l'application du modèle généré par l'apprentissage,
7. la dénormalisation des données en sortie.

9.2.3. Les avantages et les inconvénients de la méthode :

➤ **Avantages :**

Selon [37], les principaux avantages des réseaux de neurones sont :

- **Lisibilité du résultat** : le résultat de l'apprentissage est un réseau constitué de cellules organisées selon une architecture, définies par une fonction d'activation et un très grand nombre de poids à valeurs réelles.
- **Les données réelles** : les réseaux traitent facilement les données réelles "préalablement normalisées" et les algorithmes sont robustes au bruit. Ce sont, par conséquent, des outils bien adaptés pour le traitement de données complexes éventuellement bruitées comme la reconnaissance de formes (son, images sur une rétine, etc.).
- **Classification efficace** : le réseau étant construit, le calcul d'une sortie à partir d'un vecteur d'entrée est un calcul très rapide.

➤ **Inconvénients** :

Selon [37], il existe quelques limites parmi lesquelles, nous citons :

- **Temps d'apprentissage** : l'échantillon nécessaire à l'apprentissage doit être suffisamment grand et représentatif des sorties attendues. Il faut passer un grand nombre de fois tous les exemples de l'échantillon d'apprentissage avant de converger et donc le temps d'apprentissage peut être long.
- **Evolutivité dans le temps** : comme pour les arbres de décision, l'apprentissage n'est pas incrémental et, par conséquent, si les données évoluent avec le temps, il est nécessaire de relancer une phase d'apprentissage pour s'adapter à cette évolution.

9.3. La régression :

9.3.1. Définition :

La régression est un ensemble de méthodes statistiques très utilisées pour analyser la relation d'une variable par rapport à une ou plusieurs autres [52].

La régression d'une variable aléatoire y sur le vecteur de variables aléatoires x désignait la moyenne conditionnelle de y sachant x . Aujourd'hui, le terme de régression désigne tout élément de la distribution conditionnelle de y sachant x , considérée comme une fonction de x . On peut par exemple s'intéresser à la moyenne conditionnelle, à la médiane conditionnelle, au mode conditionnel, à la variance conditionnelle... [53]

En apprentissage automatique, on distingue les problèmes de régression des problèmes de classification. Ainsi, on considère que les problèmes de prédiction d'une variable quantitative sont des problèmes de régression tandis que les problèmes de prédiction d'une variable qualitative sont des problèmes de classification. Certaines méthodes, comme la régression

logistique, sont à la fois des méthodes de régression au sens où il s'agit de prédire la probabilité d'appartenir à chacune des classes et des méthodes de classification [54].

9.3.2. Principaux modèles de régression :

Il existe plusieurs modèles de régression, nous citons :

✚ Le modèle de régression linéaire :

Un modèle de régression linéaire est un modèle de régression qui cherche à établir une relation linéaire entre une variable, dite expliquée, et une ou plusieurs variables, dites explicatives [55].

✚ Le modèle de régression non linéaire :

La régression non linéaire est une méthode permettant de déterminer un modèle non linéaire de relation entre la variable dépendante et un groupe de variables indépendantes [56].

✚ Le modèle de régression quantile :

La Régression quantile est un type de régression utilisée en statistiques. Alors que la méthode des moindres carrés fournit une estimation de la moyenne conditionnelle de la variable réponse étant donné certaines valeurs des variables prédictives, la régression quantile donne par approximation soit la médiane soit les autres quantiles de la variable réponse [57].

✚ Le modèle de régression logistique :

La régression logistique est l'un des modèles d'analyse multivariée les plus couramment utilisés en épidémiologie. Elle permet de mesurer l'association entre la survenue d'un événement (variable expliquée qualitative) et les facteurs susceptibles de l'influencer (variables explicatives) [58].

9.3.3. Les avantages et les inconvénients de la méthode :

➤ Avantages :

Selon [67], les principaux avantages de la régression sont :

- Assez simple d'utilisation.
- Spécialement adapté pour les variables discrètes.
- Fournir la flexibilité dans la modélisation.

➤ **Inconvénients :**

Selon [68] et [69], il existe quelques limites parmi lesquelles, nous citons :

- Identifier les variables indépendantes.
- Les variables des résultats limités.
- Observations indépendante requise.
- Plus on complexifie le modèle, plus la variabilité résiduelle peut être – apparemment- expliquée.
- Les conditions de linéarité étant souvent précaires et limitées à un domaine de X , l'extrapolation des résultats est toujours hasardeuse.
- La régression produit un modèle interpole bien les points mais ne possède aucune valeur d'extrapolation.

9.4. Les classifieurs (SVM) :

9.4.1. Définition :

Les machines à vecteurs de support (Support Vector Machine, SVM) appelés aussi séparateurs à vaste marge sont des techniques d'apprentissage supervisées destinées à résoudre des problèmes de classification. Les SVM sont une généralisation des classifieurs linéaires [59].

9.4.2. Principe de la technique SVM

Cette technique est une méthode de classification à deux classes qui tente de séparer les exemples positifs des exemples négatifs dans l'ensemble des exemples. La méthode cherche alors l'hyperplan qui sépare les exemples positifs des exemples négatifs, en garantissant que la marge entre le plus proche des positifs et des négatifs soit maximale. Cela garantit une généralisation du principe car de nouveaux exemples pourront ne pas être trop similaires à ceux utilisés pour trouver l'hyperplan mais être situés d'un côté ou l'autre de la frontière. L'intérêt de cette méthode est la sélection de vecteurs supports qui représentent les vecteurs discriminant grâce auxquels est déterminé l'hyperplan. Les exemples utilisés lors de la recherche de l'hyperplan ne sont alors plus utiles et seuls ces vecteurs supports sont utilisés pour classer un nouveau cas, ce qui peut être considéré comme un avantage pour cette méthode [60].

9.4.3. Classifieur linéaire :

Un classifieur est dit linéaire lorsqu'il est possible d'exprimer sa fonction de décision par une fonction linéaire en x . On peut exprimer une telle fonction par [60] :

$$h(x) = \langle w, x \rangle + b = \sum_{i=1}^n w_i x_i + b$$

Où $w (\in \mathbb{R}^n)$ est le vecteur de poids et $b (\in \mathbb{R}^0)$ le biais, alors que x est la variable du problème. X est l'espace d'entrée et qui correspond à \mathbb{R}^n , où n est le nombre de composantes des vecteurs contenant les données. Notons que l'opérateur $\langle \rangle$ désigne le produit scalaire usuel dans \mathbb{R}^n . w et b sont les paramètres à estimer de la fonction de décision $h(x)$.

Pour décider à quelle catégorie un exemple estimé x' appartient, il suffit de prendre le signe de la fonction de décision : $y = \text{sign}(h(x'))$. La fonction $\text{sign}(\cdot)$ est appelée classifieur. Géométriquement (voir figure), cela revient à considérer un hyperplan qui est le lieu des points x satisfaisant $\langle w, x \rangle + b = 0$. En orientant l'hyperplan, la règle de décision correspond à observer de quel côté de l'hyperplan se trouve l'exemple x' .

On voit que le vecteur w définit la pente de l'hyperplan (w est perpendiculaire à l'hyperplan). Le terme b quant à lui permet de translater l'hyperplan parallèlement à lui-même.

L'objectif de la discrimination linéaire est de trouver la bonne fonction de décision $h(x)$. La classe de tous les hyperplans qui en découle sera notée H .

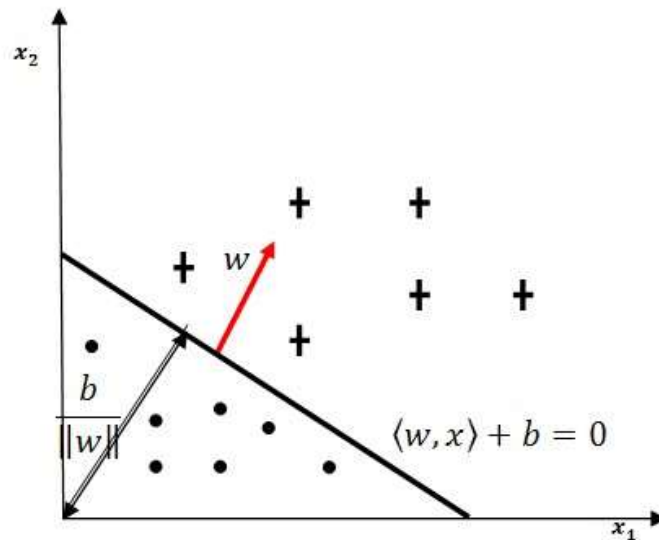


Figure 08 : hyperplan séparateur

9.4.4. Les avantages et les inconvénients de la méthode :

➤ **Avantages :**

Selon [70], les principaux avantages de SVM sont :

- Résultats en général équivalents et souvent meilleurs.
- La dérivation et l'implémentation sont simples.
- SVM est une méthode particulièrement bien adaptée pour traiter des données de très haute dimension.

➤ **Inconvénients :**

Selon [70], il existe quelques limites parmi lesquelles, nous citons :

- **Problèmes i.i.d** : (données indépendantes et identiquement distribuées)
- Grande quantité d'exemples en entrées implique un calcul matriciel important.
- Temps de calcul élevé lors d'une régularisation des paramètres de la fonction noyau.

10. La prédiction basée sur le raisonnement à base de cas

La prédiction basée sur le raisonnement à base de cas s'appuie sur des modèles (patterns) représentant des cas, qui ont été élaborés antérieurement lors d'une analyse descriptive des données faisant l'objet de la prédiction. Ces modèles sont enregistrés dans des bases de données (base de cas).

10.1. Le raisonnement à base de cas :

Le raisonnement à partir de cas est un type de raisonnement qui copie le comportement humain qui consiste à faire naturellement appel à l'expérience pour résoudre les problèmes de la vie quotidienne, en se remémorant les situations semblables déjà rencontrées et en les comparant à la situation actuelle pour construire une nouvelle solution qui, à son tour, s'ajoutera à l'expérience.

Ce type de raisonnement résout les problèmes en retrouvant des cas analogues dans sa base de connaissances et en les adaptant au cas considéré [38].

10.2. Les étapes du processus du raisonnement à base de cas :

Un système RàPC dispose d'une base de cas. Chaque cas possède une description et une solution. Pour utiliser ces informations, un moteur est aussi présent. Celui-ci va retrouver les cas similaires au problème posé. Après analyse, le moteur fournit une solution adaptée qui doit être validée. Enfin le moteur ajoute le problème et sa solution dans la base de cas [39].

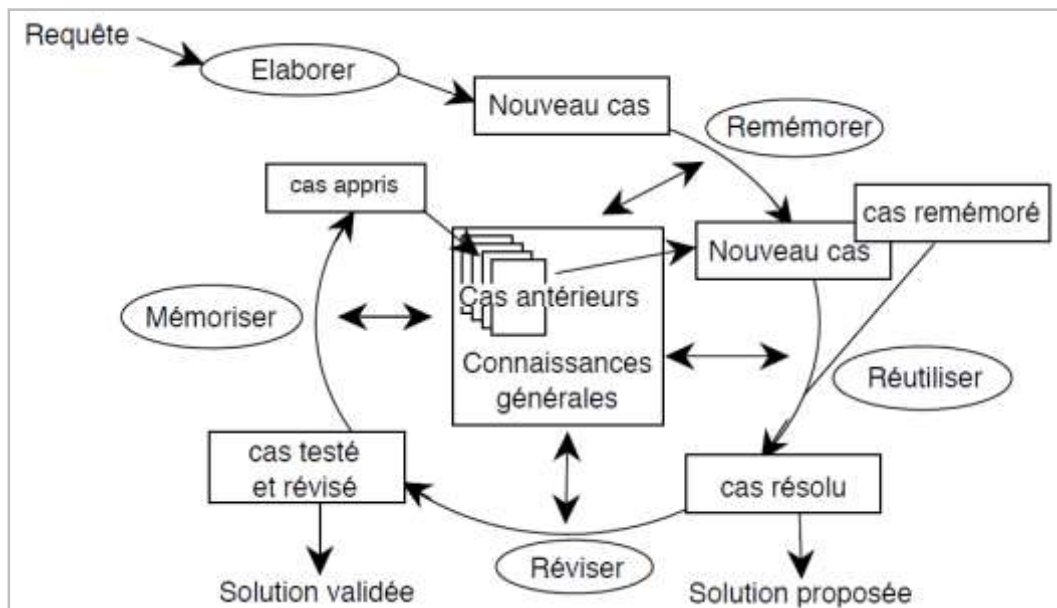


Figure 09 : les principales étapes dans le processus d'un système de raisonnement par cas.

Ce schéma présente bien les principales étapes dans le processus d'un système de raisonnement par cas. De ces étapes se dégagent trois problèmes majeurs [40, 41] :

- **L'élaboration** : à partir d'une requête, il s'agit de construire un nouveau cas et de collecter les informations permettant de formuler le problème à résoudre.
- **La remémoration** : consiste à chercher dans la mémoire un cas passé résolu. Il s'agit d'identifier les caractéristiques les plus pertinentes, les indices, et de les utiliser pour rechercher des cas passés potentiellement réutilisables. Les cas de la mémoire sont ensuite appariés avec le problème courant afin d'évaluer leur similarité avec le problème courant, et de les ordonner par similarité croissante. Le critère de similarité est défini de manière à sélectionner le cas qui aura des chances d'être adapté le plus facilement.
- **La réutilisation** : la solution du nouveau cas est déterminée en copiant et éventuellement en adaptant la solution du cas remémoré pour tenir compte des différences entre les spécifications des problèmes. La solution obtenue est proposée pour le nouveau problème.
- **La révision** : la solution du nouveau problème obtenue par adaptation est proposée. Les retours sur le succès ou l'échec résultant de la proposition permettent de réaliser des corrections. La révision permet d'identifier d'éventuelles causes d'échecs, et de proposer des adaptations supplémentaires en fonction d'une analyse des échecs. Cette étape est à la base du processus d'apprentissage pour l'amélioration des connaissances.
- **La mémorisation** : le nouveau cas est intégré dans la mémoire des cas passés. Cette intégration doit tenir compte des caractéristiques pertinentes par mise en place d'index appropriés.

La mémoire des cas est ainsi enrichie. L'apprentissage consiste à mémoriser le cas résolu et à réaliser des généralisations à partir des cas pour générer de nouvelles connaissances.

10.3. **Avantages :**

Selon [37], les principaux avantages du raisonnement à base de cas sont :

- Le raisonnement dans des domaines qui n'ont pas été bien définis et modélisés.
- La réduction des tâches d'acquisitions des connaissances.
- L'évitement des erreurs répétitives produites dans le passé.
- Fournir la flexibilité dans la modélisation des connaissances.
- L'amélioration de la qualité des solutions.
- L'apprentissage sans cesse : Les systèmes RàPC peuvent apprendre et s'enrichir par l'acquisition de nouvelles connaissances de cas.
- Fournir un moyen argumentaire : Les systèmes du raisonnement à partir de cas peuvent fournir un cas antérieur et sa solution pour aider à convaincre un utilisateur, ou à justifier, une solution proposée au problème en cours.
- Le reflet de raisonnement humain : Il y a beaucoup de situations où nous – les humains - utilisons une forme de raisonnement à partir de cas.

11. **Conclusion**

Dans ce chapitre nous avons expliqué les différents concepts relatifs aux domaines de la prédiction des parcours d'apprentissage (le modèle prédictif, les méthodes de la prédiction, des exemples des méthodes prédictives) et plus précisément la méthode de raisonnement à base des cas que nous allons appliquer par la suite dans notre contribution.

Dans le chapitre suivant, nous allons expliquer la conception de notre système en proposant une architecture de la prédiction des parcours d'apprentissage basée sur les fichiers logs et la méthode de raisonnement à base de cas dans le système E-learning.

Chapitre 3 : Conception et architecture

7. Introduction :

Dans ce chapitre, nous allons présenter la conception de notre système en proposant une architecture et une démarche pour la prédiction des parcours d'apprentissage à partir des fichiers logs générés automatiquement par les systèmes d'E-learning.

8. Problématique et Objectif :

Lors de l'apprentissage à distance et avec les systèmes d'E-learning, les formateurs, les enseignants ou les pédagogues, par manque de contact face à face, ne peuvent pas bien encadrer les apprenants et suivre leur parcours d'apprentissage réel et voir s'il est conforme ou non au scénario pédagogique préétabli au préalable et vérifier si ces parcours entrepris par les apprenants mèneront à une réussite ou à un échec d'apprentissage. Le seul moyen qui existe actuellement, pour valider le succès de l'apprentissage d'un cours est en fin de la formation par un examen final. C'est-à-dire, l'enseignant doit attendre la fin de la formation pour valider la bonne assimilation de son cours par ses apprenants et cela n'est pas très efficace parce que ça sera un peu trop tard pour remédier aux problèmes de l'apprentissage, probablement rencontrés par leurs apprenants lors des différentes sessions d'apprentissage, favorisant ainsi leur désengagement et échec.

En effet, et par manque de connaissances sur les comportements de ses apprenants, l'enseignant ne peut pas intervenir en juste temps pour proposer des aides et des recommandations pour les apprenants en difficultés d'apprentissage.

Pour rendre les systèmes de E-learning plus performant et garantir un apprentissage réussi, nous proposons de leur intégrer un outil d'aide à la décision dédié à l'analyse et la prédiction des parcours d'apprentissage des apprenants en cours d'apprentissage, qui permettra aux formateurs que ce soit des enseignants, des tuteurs ou des pédagogues, de suivre l'apprentissage de leurs apprenants et prédire leurs parcours d'apprentissage futurs permettant ainsi l'évaluation précoce de leur comportement d'apprentissage, s'il s'agira d'un comportement positive favorisant la réussite ou négatif amenant à un échec.

Notre travail se focalise plus sur la tâche de prédiction qui complète une tâche d'analyse des parcours déjà réalisée.

En fait, la prédiction des parcours des apprenants en cours d'apprentissage est basée sur les résultats obtenus des analyses des parcours d'apprentissage et plus particulièrement sur les modèles des parcours d'apprentissage découverts à partir de l'historique d'apprentissage (fichiers log) relatif aux apprenants, inscrits au même cours, ayant terminé leur apprentissage.

Exigences Fonctionnelles

Notre Système dédié à la prédiction des parcours d'apprentissage baptisé *Trails_predictor*, offre les fonctionnalités suivantes :

- Pré-traitement des fichiers log (nettoyage et filtrage),
- Identifier le parcours d'apprentissage de chaque apprenant à partir du fichier log.
- Afficher les parcours d'apprentissage des apprenants en cours d'apprentissage sous forme textuelle.
- Dédire les modèles des parcours d'apprentissage à partir des résultats de l'analyse des parcours d'apprentissage et plus précisément, à partir des clusters des apprenants ayant eu un comportement d'apprentissage similaire.
- Prédire les parcours d'apprentissage futurs des nouveaux apprenants en cours d'apprentissage du même cours en se basant sur les clusters d'actions obtenus lors de la phase d'analyse.
- Evaluer la nature positive (réussite) ou négative (échec d'apprentissage) des comportements d'apprentissage relatifs aux parcours prédits des apprenants en les comparant avec les modèles des parcours déduits ultérieurement. Cette évaluation est basée sur un raisonnement à base de cas (modèles des parcours).

9. Le contexte global du système Trails_predictor

Notre système baptisé *Trails_predictor* est utilisé par les responsables des formations à savoir : les enseignants, les formateurs, les pédagogues ou les tuteurs pour extraire des connaissances sur le processus d'apprentissage et prédire d'une part, les parcours futurs des apprenants en cours d'apprentissage et d'autre part la réussite ou l'échec de leur apprentissage anticipant ainsi les situations problématiques.

Trails_predictor utilise comme entrée, d'une part, les fichiers logs qui représentent les journaux de l'historique de l'apprentissage en cours générés automatiquement par les plateformes d'apprentissage et d'autre part, les résultats normalement d'un autre système dédié à l'analyse descriptive des parcours d'apprentissage, pour extraire des connaissances sur l'apprentissage

en cours et prédire le parcours d'apprentissage à entreprendre par chaque apprenant ainsi que le résultat de son apprentissage (réussite ou échec).

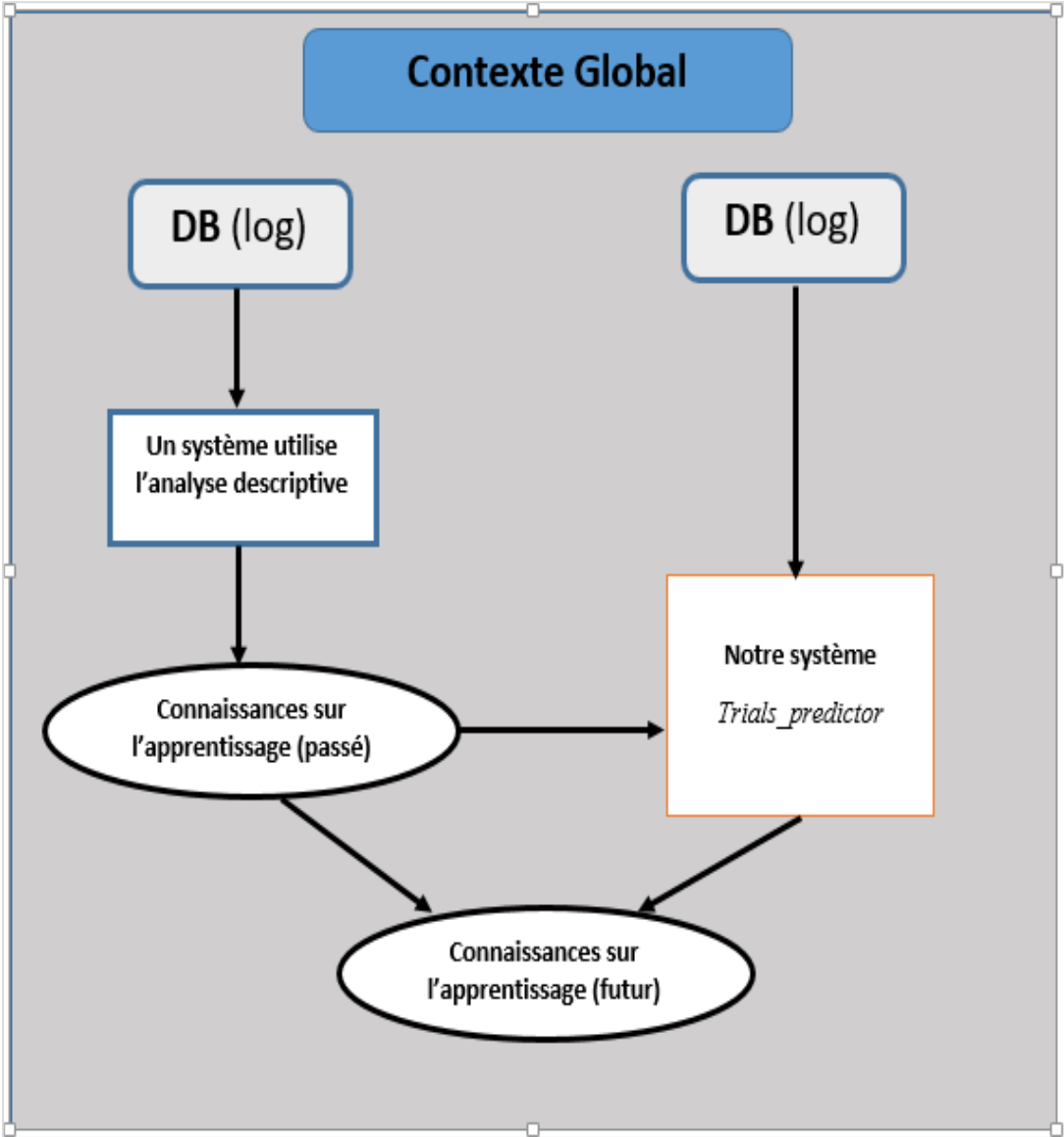


Figure 10 : Contexte Global du système Trails_prédicator

10.L'Architecture du Système Trails_predictor

L'architecture du système *Trails_predictor* se base sur trois modules :

- Un module de prétraitement de l'historique de l'apprentissage qui traite, nettoie et filtre les fichiers log.
- Un module d'extraction des parcours des apprenants en cours d'apprentissage.
- Un module de déduction des modèles des parcours à partir des résultats de l'analyse.
- Un module de prédiction des parcours, qui se base sur les parcours extraits et les clusters d'actions de l'analyse pour formuler les parcours prédits.
- Un module évaluation des parcours prédits, qui se base sur les parcours prédits et le modèle des parcours d'apprentissage.

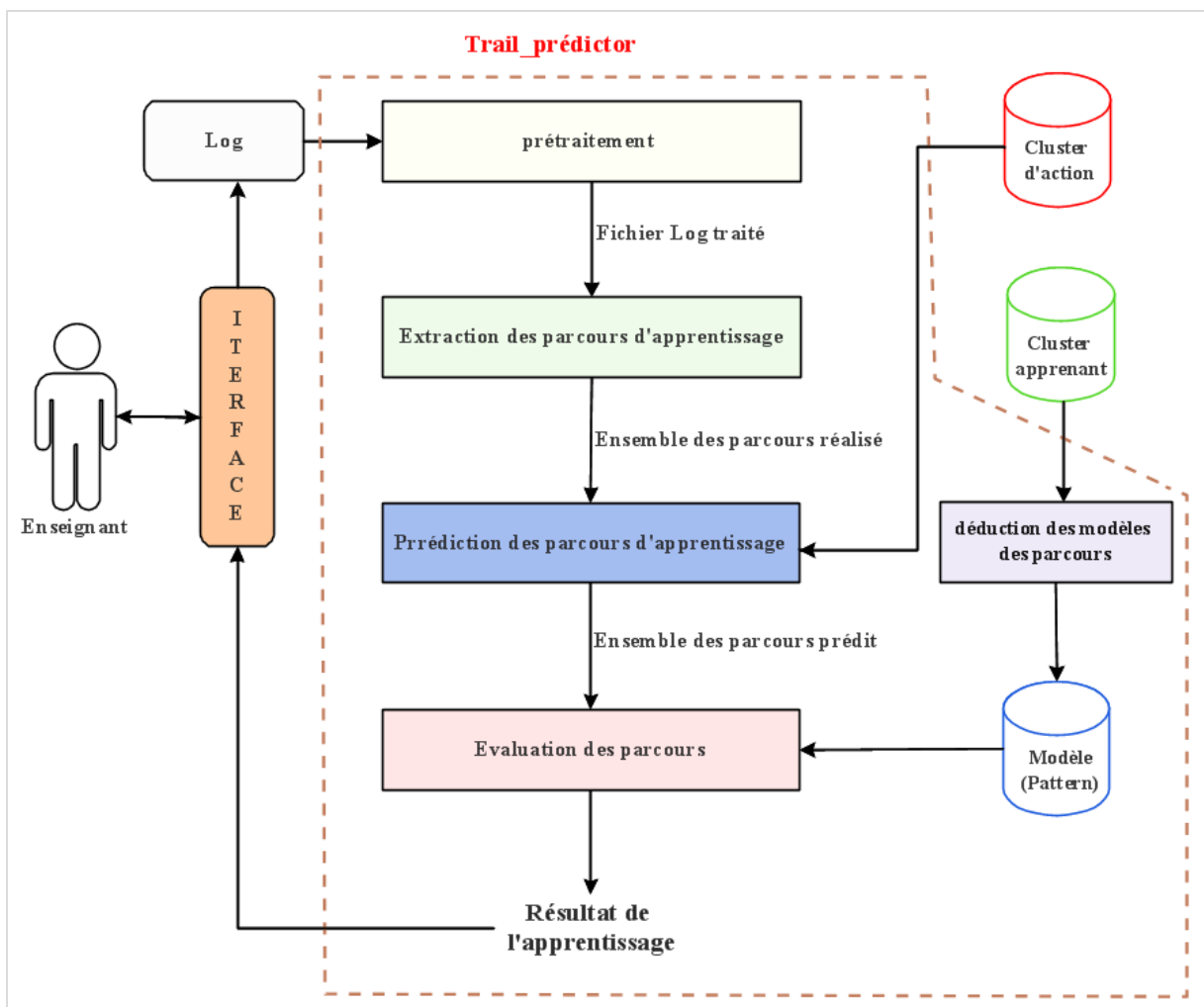


Figure 11 : L'Architecture du système Trails_predictor

10.1. Fichier LOG :

Le fichier Log est un fichier représentant l'historique de l'apprentissage qui contient une liste des observés qui sont automatiquement générés par le système d'apprentissage en ligne lors des différentes interactions des utilisateurs avec le système. Chaque observé est constitué principalement des attributs relatifs à la date et l'heure de la collecte, de l'identifiant de l'utilisateur (nom ou adresse IP), de l'action effectuée. Les observés peuvent contenir d'autres attributs. Il est à noter que le fichier log enregistre tous les observés relatifs aux différentes interactions de tous les utilisateurs (apprenants, enseignants, administrateur) avec le système d'apprentissage.

10.2. Le module de Traitement log :

L'objectif de ce module est d'éliminer le bruit et les erreurs générées par le serveur et de sélectionner les observés relatifs seulement aux apprenants qui sont inscrits à un cours donné faisant l'objet de l'étude. Le module de *traitement Log* est basé sur deux étapes consécutives à savoir : le nettoyage et le filtrage.

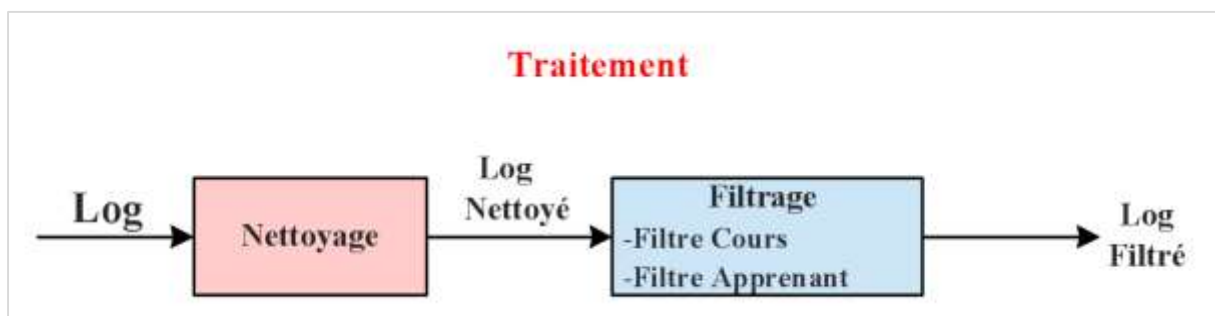


Figure 12 : Module de « Traitement Log »

10.2.1. Nettoyage :

Dans cette étape, il s'agit d'éliminer le bruit (par exemple les actions qui se répètent successivement) et de supprimer les erreurs retournées par le serveur quand on veut afficher un contenu qui n'existe pas, un accès non autorisé ou une expiration de délais ...etc.

10.2.2. Filtrage :

Dans cette étape, il s'agit de sélectionner seulement les observés relatifs aux utilisateurs de type apprenants, inscrits à un cours donné. Pour ce faire, on applique au fichier log deux types de filtres : le *filtre cours* puis le *filtre apprenant*.

- Le *filtre cours* filtre et supprime du fichier log, les observés relatifs aux cours non faisant l'objet de l'étude pour avoir seulement les observés essentiels.
- Le *filtre apprenant* sélectionne du fichier log filtré, seulement les observés relatifs aux utilisateurs de type apprenant.

10.3. Le module extraction des parcours d'apprentissage :

Le module de prétraitement de *Trails_predictor*, génère comme résultat : un fichier log traité qui ne contient pas les bruits et les erreurs générées par le serveur.

Ce fichier log contient seulement les observés relatifs aux apprenants qui sont inscrits à un cours donné faisant l'objet de l'étude.

Ce log contient tous les parcours d'apprentissage des apprenants en cours d'apprentissage sous forme textuelle et qui sont affichés par notre système *Trails_predictor*.

10.4. Le module de génération des modèles des parcours :

Un système utilise l'analyse descriptive génère comme résultat : les clusters relatifs aux apprenants ayant eu le même comportement et les clusters des actions fortement corrélées.

Le module de *génération des modèles des parcours* peut se baser sur le résultat d'un système d'analyse et plus particulièrement sur les clusters des apprenants. Cette génération des modèles peut passer par deux étapes par exemple :

1. Déduction d'un modèle de parcours pour chaque cluster groupant les parcours des apprenants ayant eu un comportement similaire. Cette déduction est basée sur une comparaison entre les parcours des apprenants de chaque cluster, pour en extraire le chemin le plus fréquent sinon le plus long.

Fonction Modèle (L)

Paramètre :

L : Liste des parcours du cluster de m éléments (actions)

Variable :

A ← liste vide ;

Début

Pour chaque cluster faire

Répéter

Comparer le parcours de l'apprenant (i) **avec** le parcours de l'apprenant (j)

Si (le parcours de l'apprenant (i)= le parcours de l'apprenant (j)) **alors**

Enregistrer le parcours dans la liste A

Sinon

Passer à l'apprenant suivant ;

Fin si

Jusque (la fin de la liste L)

Sortie : Liste A

Fin

Table 02 : Algorithme pour extraire le modèle de chaque cluster

2. Etiquetage des modèles des parcours déduits par une valeur indiquant soit la réussite de l'apprentissage (+1) soit l'échec (-1) et cela en se basant sur l'historique de l'apprentissage. Cet étiquetage est fait manuellement par l'enseignant.

10.5. **Le module de prédiction des parcours d'apprentissage :**

La prédiction des parcours d'apprentissage se base d'une part, sur les parcours des apprenants en cours d'apprentissage qui ont été extraits antérieurement par le module *d'extraction des parcours d'apprentissage* et d'autre part, sur les clusters des actions générés par une analyse descriptive.

En fait, La prédiction des parcours d'apprentissage utilise les couples d'actions qui définissent les transitions possibles entre les différentes actions réalisées par les apprenants.

Le système réalise une prédiction pour chaque parcours d'apprenant de fichier log à partir des couples d'action.

Chaque parcours complété sera un parcours prédit.

| |
|--|
| <p>Fonction prédiction (Action)</p> <p>Paramètre :</p> <p>Ch1, ch2 : chaîne de caractère ;</p> <p><u>Début</u></p> <p><i>Action</i> = la dernière action réalisée par l'apprenant</p> <p>Pour chaque cluster des apprenants faire</p> <p><i>Chercher</i> la dernière action réalisée par l'apprenant dans tous les clusters d'actions</p> <p><i>Sélectionner</i> le ou les clusters qui contiennent cette action</p> <p>Si un cluster trouvé alors</p> <p><i>Complémenter</i> le parcours par les actions fortement corrélées avec l'action réalisée par l'apprenant</p> <p>Sinon (le cas de plusieurs clusters)</p> <p>Il y a autant de parcours prédit que de clusters</p> <p>Fin si</p> <p>Sortie : le parcours prédit ;</p> <p><u>Fin</u></p> |
|--|

Table 03 : Algorithme de la prédiction des parcours d'apprentissage

10.6. Le module d'évaluation des parcours prédits :

L'évaluation de la nature positive (réussite) ou négative (échec d'apprentissage) des comportements d'apprentissage relatifs aux parcours prédits des apprenants se base sur les modèles des parcours et sur les parcours prédits.

Ce module se base sur un raisonnement à base de cas où les modèles des parcours d'apprentissage représentent la base de cas et le parcours prédit représente le nouveau cas. Le processus d'évaluation des parcours est un moteur comparant chaque parcours prédit avec la base des modèles des parcours en calculant la distance entre les deux parcours pour déduire la probabilité de la réussite ou de l'échec de chaque apprenant en cours d'apprentissage.

Fonction évaluation (model, prédit)

Paramètre :

model, prédit : chaîne de caractère représentant respectivement le modèle du parcours et le parcours prédit ;

Variable :

t, s : entier ;

Début

Action = Le $i^{\text{ème}}$ caractère dans une chaîne de caractère

Si ((la longueur de model) < (la longueur de prédit)) **alors**

Répéter

Comparer chaque action du parcours prédit avec son correspondant dans le parcours modèle

Si les deux actions sont similaires **alors**

Comparer les actions suivantes ;

Incrémenter la variable s relative au taux de similarité

Fin si

Jusqu'à la fin de la chaîne modèle

// $t \leftarrow \text{mod} * 100 / \text{la longueur de prédit}$ //

Calculer le pourcentage de simulation entre les deux chaînes de caractères

Affecter le résultat dans la variable t.

| |
|---|
| <p>Sinon</p> <p>Répéter</p> <p>Si les deux actions sont similaires alors</p> <p><i>Comparer</i> les actions suivantes ;</p> <p><i>Incrémenter</i> la variable s relative au taux de similarité</p> <p>Fin si</p> <p>Jusqu'à la fin de la chaîne prédit</p> <p>//t← prd*100/la longueur de model//</p> <p><i>Calculer</i> le pourcentage de simulation entre les deux chaînes de caractères</p> <p><i>Affecter</i> le résultat dans la variable t.</p> <p>Fin si</p> <p>Sortie t ;</p> |
|---|

Table 04 : Algorithme de l'évaluation de la réussite

11. Conclusion

Dans ce chapitre, nous avons présenté la démarche suivie pour la prédiction des parcours des apprenants en cours d'apprentissage ainsi que l'architecture et le fonctionnement de notre système *Trails_predictor* dédié à la prédiction des parcours d'apprentissage à partir des fichiers logs générés automatiquement par les systèmes d'E-learning. Cette prédiction est basée sur les modèles des parcours d'apprentissage créés lors d'une analyse descriptive de l'historique de l'apprentissage.

Dans le chapitre suivant, nous allons présenter notre système et expliquer comment nous avons l'implémenté.

Chapitre 4 : Implémentation

3. Introduction

Ce chapitre sera consacré à la mise en pratique des idées conceptuelles évoquées au niveau du chapitre précédent. Nous commençons ce chapitre par la description de la plate-forme matérielle et logicielle et nous le terminons par la présentation de notre application.

4. Description des plates-formes utilisées

Un type de plate-forme est à présenter ici, au niveau de la plate-forme matérielle, on va présenter la machine auquel on a réalisé et tester notre Système, avec une description de la configuration matérielle de l'ordinateur utilisée pendant la phase de programmation et du test de l'application. Une plate-forme logicielle représente les outils, logicielles et les langages de programmation orientés web utilisés, qui sont nécessaires à la réalisation, avec une description de chacun d'eux et motivation de choix de ses outils et langages.

4.1. Plate-forme matérielle

Pour réaliser l'application, et pour la tester, on a utilisé les machines suivantes :

➤ **PC** : pc (portable laptop)

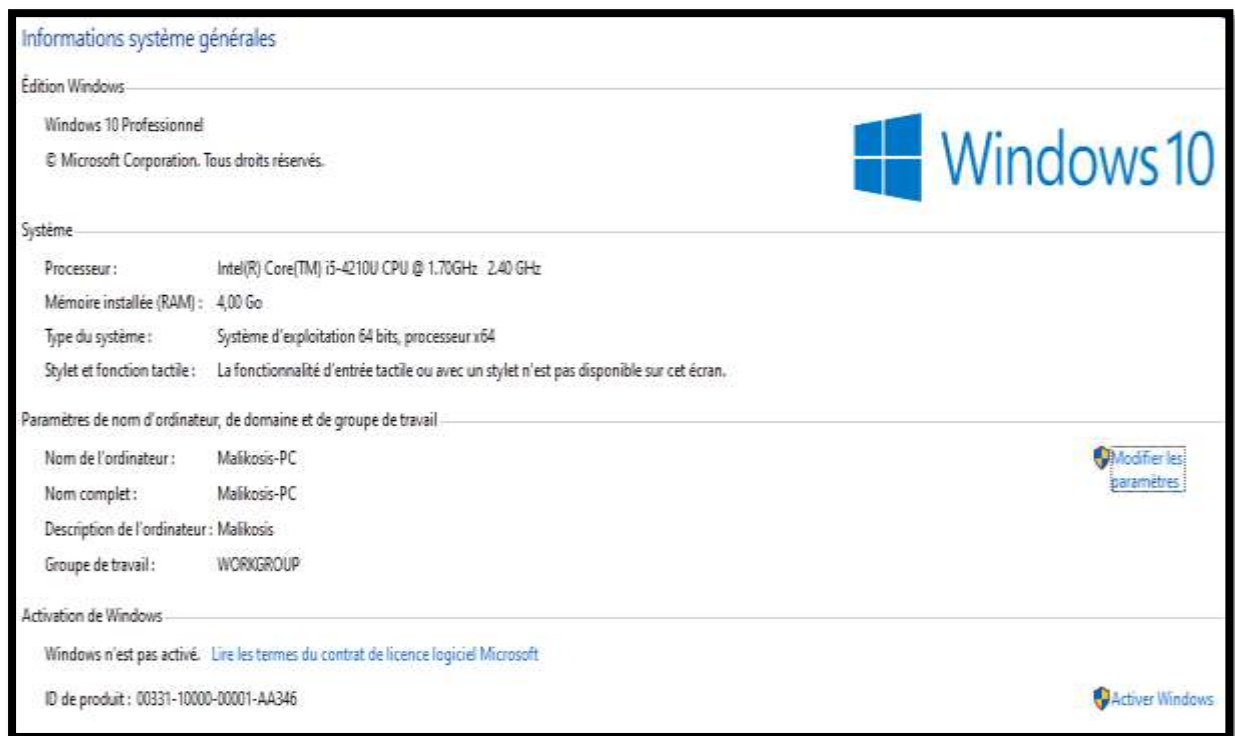


Figure 13 : Informations de la machine (PC)

Configuration matérielle de la machine de développement

| | |
|------------------------|------------------------|
| Système d'exploitation | Windows |
| Type de systèmes | 64bits |
| Processeur | Intel I5 1.7Ghz 2.4Ghz |
| Mémoire de RAM | 4 GO |

Table 05 : Caractéristiques du système d'exploitation

4.2. Plate-forme logicielle

4.2.1. Système d'exploitation

L'application a été testée sur les systèmes suivants :

| SE | Version | Caractéristiques |
|--------------|---------|--------------------------|
| Windows (PC) | 10 | Windows 10 Professionnel |

Table 06 : Caractéristiques du système d'exploitation.

4.2.1.1. Windows 10 :

Windows 10 est un système d'exploitation de la famille Windows NT développé par la société américaine Microsoft. Officiellement présenté le 30 septembre 2014, il est disponible publiquement depuis le 29 juillet 2015. Bien que le système s'appelle Windows 10, il s'agit de la version NT 6.4 pour les versions jusqu'à la « Technical Preview », la première version de Windows NT 6 étant Windows Vista. Windows 10 est ainsi une ultime version de Windows NT 6.0 ; néanmoins, depuis la version finale, il porte bel et bien le numéro interne 10 en lieu et place de 6.4. Il est le successeur de Windows 8.1. [66]

4.2.2. Outils de développement

| Outil | Version | Description |
|------------------------|---------|--|
| Apache NetBeans IDE | 12.1 | créer des programmes dans de nombreux langages de programmation (Java, C++, PHP...). |

Table 07 : les outils de développement

2.2.2.1. Environnement de développement NetBeans :

L'environnement de développement utilisé, est NetBeans 12.1, il possède de nombreux Avantages qui sont à l'origine de son énorme succès dont les principaux sont :

- Un environnement de développement intégré (IDE).
- Permet de supporter différents autres langages, comme Python, C, C++, JavaScript, XML, Ruby, PHP 8 et HTML. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages Web).
- La construction incrémentale des projets JAVA grâce à son propre compilateur, qui permet en plus de compiler le code même avec des erreurs, de générer des messages d'erreurs personnalisés, de sélectionner la cible, ...

2.2.2. Langages utilisés

| Langage | Version | Description |
|---------|---------|--|
| Java | 15.0.2 | Java est un langage de programmation orienté objet |

Table 08 : Langages utilisés.

Dans cette partie, on va détailler les outils, logicielles et les langages de programmation utilisés dans la réalisation de notre application :

4.2.3.1. Le langage JAVA

C'est un langage de programmation orienté objet, développé par Sun Microsystems. Il permet de créer des logiciels compatibles avec de nombreux systèmes d'exploitation (Windows, Linux, Macintosh, Solaris).

Java donne aussi la possibilité de développer des programmes pour téléphones portables et assistants personnels.

Enfin, ce langage peut être utilisé sur internet pour des petites applications intégrées à la page web (applet) ou encore comme langage serveur (jsp) [43].

➤ **La technologie JAVAFX**

JavaFX est une technologie créée par [Sun Microsystems](#) qui appartient désormais à [Oracle](#), à la suite du rachat de [Sun Microsystems](#) par [Oracle](#) le 20 avril 2009.

Avec l'apparition de Java 8 en mars 2014, JavaFX devient la bibliothèque de création d'interface graphique officielle du langage Java, pour toutes les sortes d'application (applications mobiles, applications sur poste de travail, applications Web), le développement de son prédécesseur [Swing](#) étant abandonné (sauf pour les corrections de bogues).

JavaFX est désormais une pure API Java (le langage de script spécifique qui a été un temps associé à JavaFX est maintenant abandonné).

JavaFX contient des outils très divers, notamment pour les médias audio et vidéo, le graphisme 2D et 3D, la programmation Web, la programmation multi-fils etc.

Le SDK de JavaFX étant désormais intégré au JDK standard [Java SE](#), il n'y a pas besoin de réaliser d'installation spécifique pour JavaFX.

Des projets libres complètent JavaFX en fournissant des composants absents de JavaFX proprement dit, voir par exemple JFXtras et ControlsFX [44]

5. Présentation de l'application

5.2. Page d'accueil :

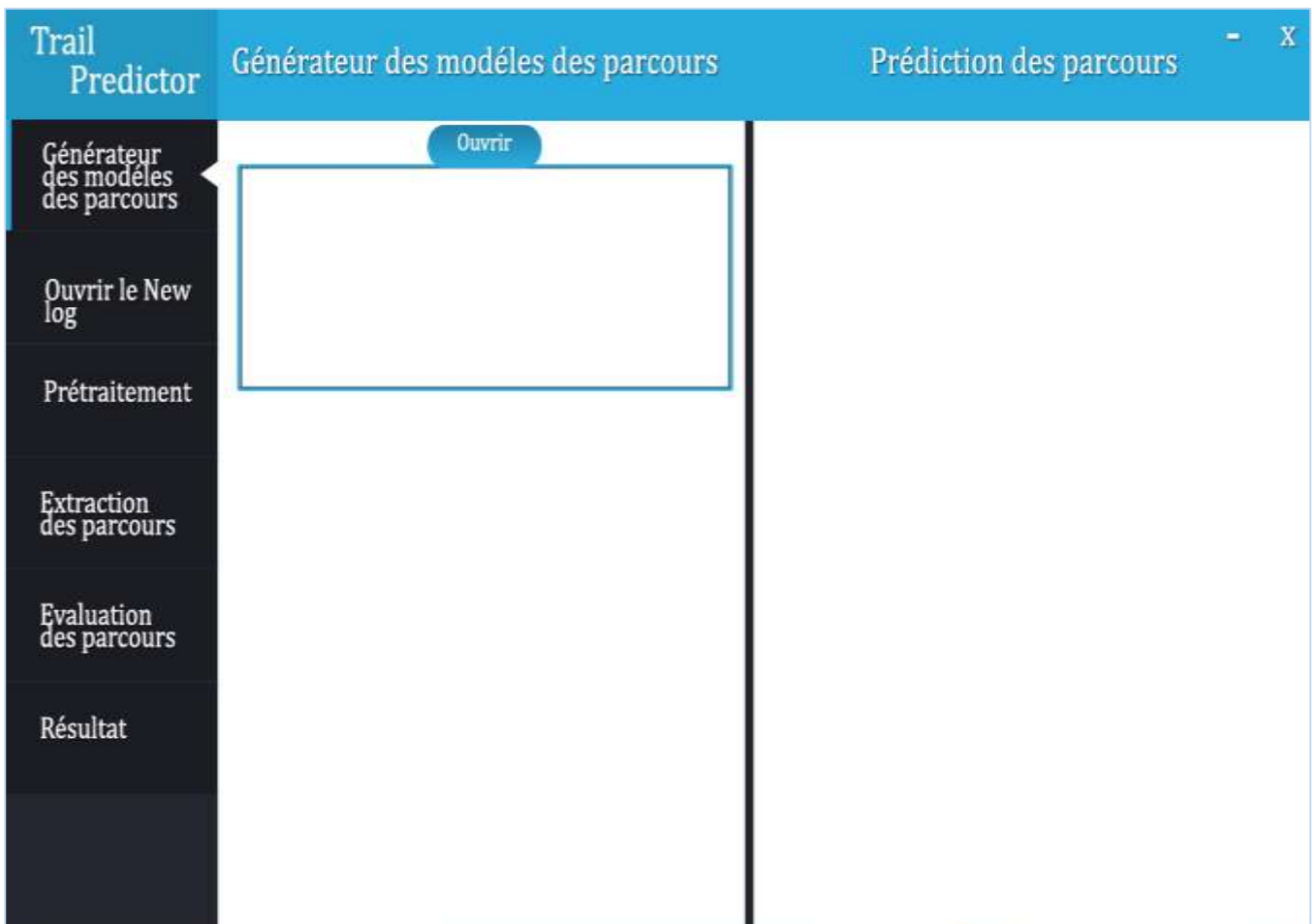
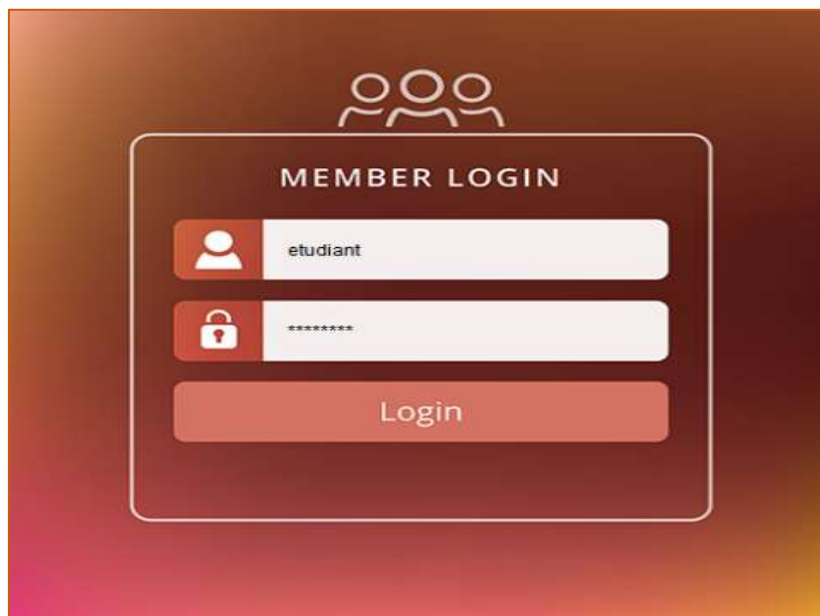


Figure 15 : Page d'accueil de l'application

5.3. S'authentifier :

L'enseignant doit s'authentifier pour accéder à son espace pour extraire des informations sur l'apprentissage de ses apprenants inscrits à ses cours.



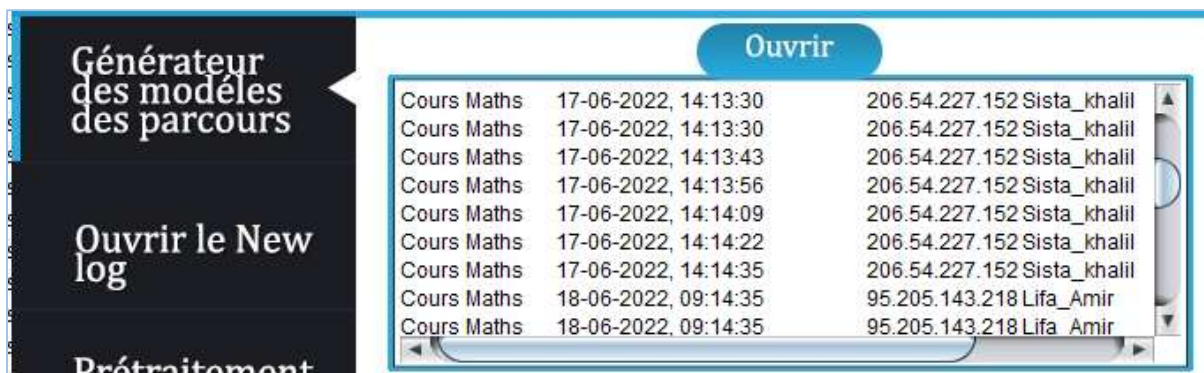
MEMBER LOGIN

etudiant

Login

Figure 16 : Formulaire d'authentification.

5.4. Ouverture du fichier log



| Cours | Date | IP | Nom |
|-------------|----------------------|----------------|--------------|
| Cours Maths | 17-06-2022, 14:13:30 | 206.54.227.152 | Sista_khalil |
| Cours Maths | 17-06-2022, 14:13:30 | 206.54.227.152 | Sista_khalil |
| Cours Maths | 17-06-2022, 14:13:43 | 206.54.227.152 | Sista_khalil |
| Cours Maths | 17-06-2022, 14:13:56 | 206.54.227.152 | Sista_khalil |
| Cours Maths | 17-06-2022, 14:14:09 | 206.54.227.152 | Sista_khalil |
| Cours Maths | 17-06-2022, 14:14:22 | 206.54.227.152 | Sista_khalil |
| Cours Maths | 17-06-2022, 14:14:35 | 206.54.227.152 | Sista_khalil |
| Cours Maths | 18-06-2022, 09:14:35 | 95.205.143.218 | Lifa_Amir |
| Cours Maths | 18-06-2022, 09:14:35 | 95.205.143.218 | Lifa_Amir |

Figure 17 : l'ouverture du fichier log

5.5. Traitement du fichier log

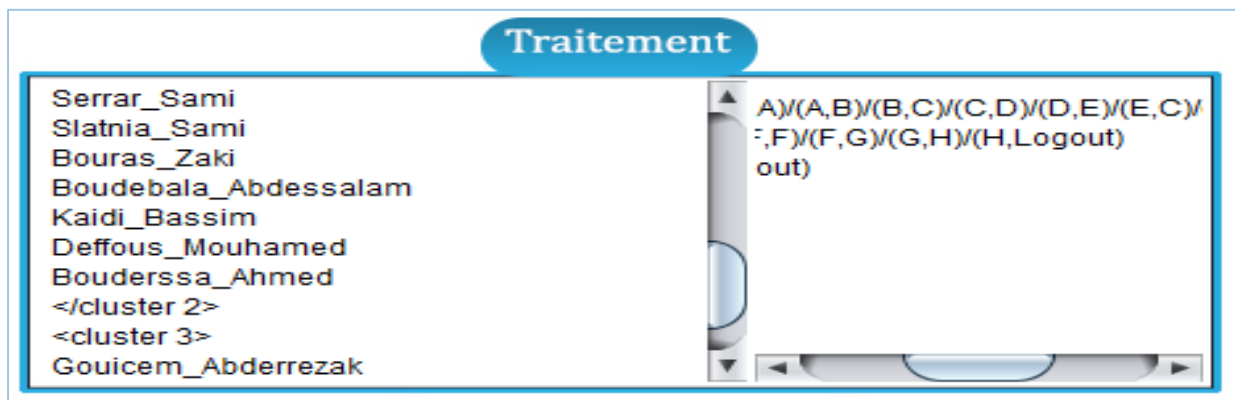


Figure 18 : Traitement du fichier log

5.6. Générer les modèles

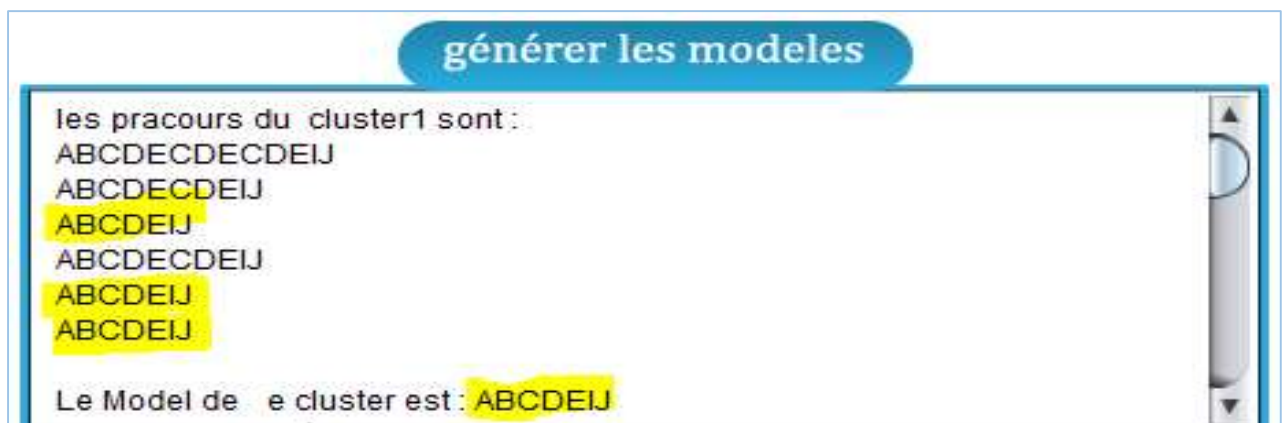


Figure 19 : Générer les modelés

5.7. Le prétraitement du fichier log



Figure 20 : prétraitement du fichier log

5.8. La prédiction des parcours d'apprentissage

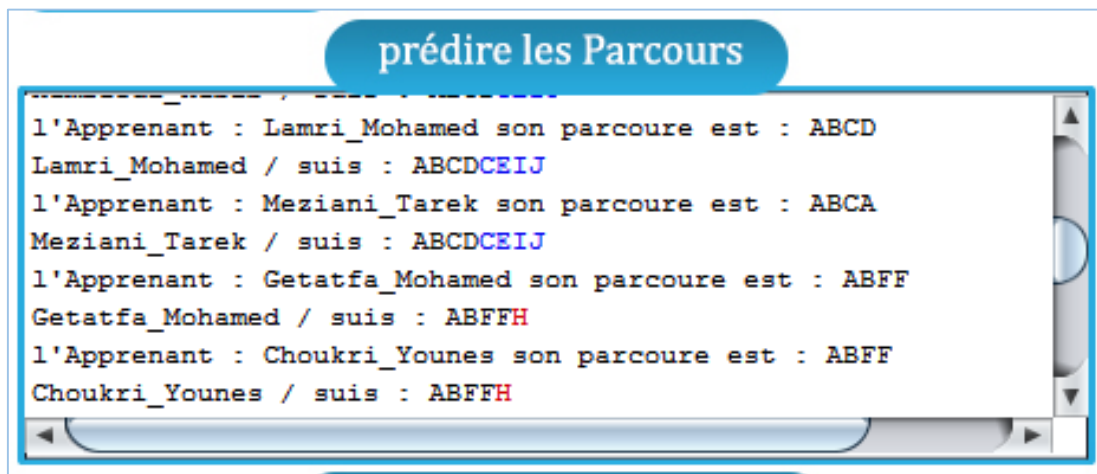


Figure 21 : Prédiction des parcours d'apprentissage

5.9. Evaluation de la réussite

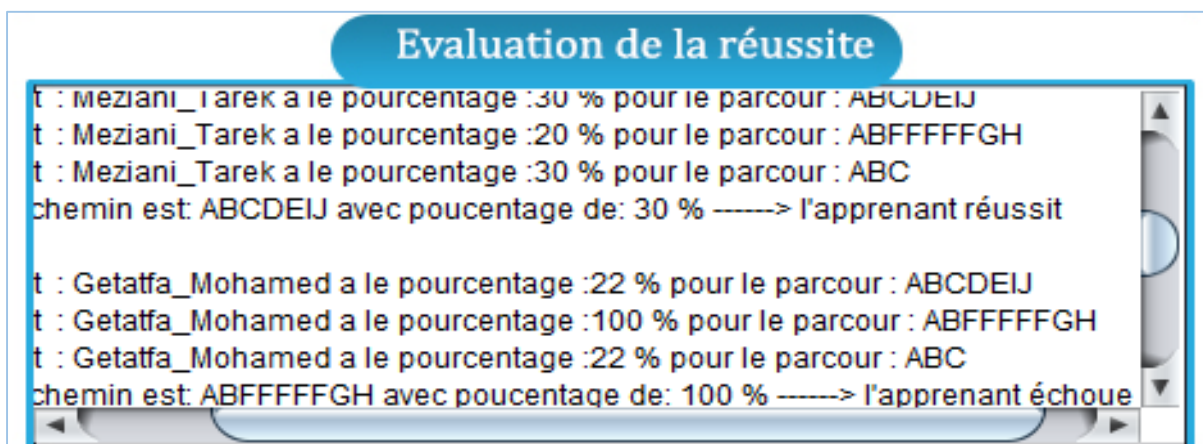


Figure 22 : Evaluation de la réussite

6. conclusion

A travers ce chapitre nous avons présenté de façon exhaustive les caractéristiques de la plate-forme matérielle et de la plate-forme logicielles, et nous avons montré la présentation de l'application afférente suivis. Nous avons arrivé à l'objectif où nous avons créé un modèle (pattern) de chaque cluster et nous avons réalisé une prédiction pour chaque parcours d'apprenant à partir des couples d'action.

Finalement nous avons évalué les nature comportements d'apprentissage relatifs aux parcours prédits des apprenants se base sur les modèles.

Conclusion générale

Dans ce manuscrit nous avons présenté notre système baptisé *Trails_predictor* qui est utilisé par les responsables des formations à savoir : les enseignants, les formateurs, les pédagogues ou les tuteurs pour prédire à partir de l'historique d'apprentissage (fichier log) d'une part, les parcours futurs des apprenants en cours d'apprentissage et d'autre part la réussite ou l'échec de leur apprentissage anticipant ainsi les situations problématiques et assurant des interventions remédiatives en juste temps. .

La prédiction est basée sur un raisonnement à base de cas qui s'appuie sur les modèles des parcours d'apprentissage représentant la base de cas. Ces modèles sont élaborés en se basant sur les résultats d'une analyse des parcours d'apprentissage. Le processus du raisonnement à base de cas se base sur une comparaison entre les parcours prédits (nouveaux cas) des apprenants en cours d'apprentissage avec les modèles des parcours d'apprentissage (base de cas).

L'évaluation du parcours prédits prédisant la réussite ou l'échec de l'apprenant en cours d'apprentissage se base sur la valeur de réussite ou non qui est étiqueté au modèle de parcours d'apprentissage sélectionné après comparaison comme étant le plus similaire au parcours prédit.

Comme perspectives, nous projetons d'intégrer notre système dans une plateforme d'E-learning tel : *moodle* pour récupérer directement le fichier log généré automatiquement par cette plateforme et lancer la prédiction. Nous projetons aussi de comparer les résultats de notre prédiction basée sur les modèles des parcours d'apprentissage découverts à partir d'une analyse descriptive des fichiers log avec les résultats d'une prédiction basée sur analyse prédictive qui s'appuie sur les algorithmes d'apprentissage automatique (machine learning) tels : les arbres de décision, les réseaux de neurones ou les SVMs.

Références

- [1] <http://tecfa.unige.ch/tecfa/research/pnr33/report/nti-ch-27.html>
- [2] <http://lesdefinitions.fr/enseignement>
- [3] <http://lestroiscouronnes.esmeree.fr/didactique-information/l-objet-d-etude-des-didactiques-et-leurs-trois-heuristiques>
- [4] <http://www.fffod.org/media/2002-baba-foad.pdf>
- [5] <http://etudiant.aujourd'hui.fr/etudiant/info/formation-a-distance-cours-par-correspondance.html>
- [6] https://fr.wikipedia.org/wiki/Cours_par_correspondance
- [7] https://fr.wikipedia.org/wiki/Formation_%C3%A0_distance
- [8] S. BELLIER, Le e-Learning : pédagogie contenus modalités acteurs, Collection Entreprise & Carrières, Editions Liaisons, Cegos, Rueil-Malmaison, 2001.
- [9] Iness NEDJI MILAT, «L'identification de l'émotion des apprenants dans les Plates-formes de E-learning », Mémoire de diplôme de MAGISTER, UNIVERSITE BADJI MOKHTARANNABA, Faculté des sciences de l'ingénieur, Année : 2008.
- [10] <http://www.elearningagency.com>
- [11] F. Maamar, M. Izzeddine, «Réalisation d'une application visioconférence dédiée à la téléformation des professionnels de santé», Université Abou Bakr Belkaïd de Tlemcen, Année MAI 2016.
- [12] Amor BOULEDROUA, « *Découverte des préférences des apprenants dans un environnement de e-Learning* », Mémoire de diplôme de MAGISTER, UNIVERSITE BADJI MOKHTARANNABA, Faculté des sciences de l'ingénieur, Année : 2012.
- [13] <https://revolution-rh.com/e-learning-avantages-inconvenients/>
- [14] Dossou Anani Koffi DOGBE-SEMANOU., Anne Durand., "*Etude comparative de plates-formes de formation à distance*", dans le cadre du Projet @2L, avril 2008.
- [15] Hathaway, T., et al. (2007). "Report on pedagogical practices and methods in e-learning." School of Education, University of Wales, Bangor. Retrieved May 10: 2009.
- [16] Marwa BEKRAR, «Protection de la vie privée à base d'agents dans un système d'e-learning », Mémoire de diplôme Ingénieur d'Etat, Ecole nationale Supérieur d'Informatique, Année : 2013.
- [17] Finkelstein, J. (2006). Learning in real time. San Francisco, CA: Jossey-Bass. Page58.

- [18] Daly Cris, (2009) "A Learning Management System ... or an Intelligent Tutoring System?" <http://www.learningsolutionsmag.com/articles/45/learning-management-system-or-an-intelligent-tutoring-system>
- [19] <https://fr.wikipedia.org/wiki/E-learning>
- [20] H.Chorfi, M.Jemni, télé-enseignements, état de l'art et perspectives, JTEA 2002, 21-22-23 mars 2002, Sousse Nord, Tunisie
- [21] <https://www.1min30.com/dictionnaire-du-web/fichier-log-definition>
- [22] <http://www.dicodunet.com/definitions/hebergement/fichier-log.htm>
- [23] SETTOUTI Lotfi-Sofiane, PRIE Y, MARTY Jean-Charles, MILLE Alain. Vers des systèmes à base de traces modélisées pour les E.I.A.H. Workshop Traces, interactions, co-constructions collectives et relations à la cognition, AS CoMETE, Paris, Avril 2007.
- [24] David et al., 2005
- [25] El-Golli.A, Rossi.F, Conan-Guez.B, et Lechevallier.Y (2006). Une adaptation des cartes auto-organisatrices pour des données décrites par un tableau de dissimilarités. Revue de statistique Appliquée, LIV(3) :33-64
- [26] Robert Nadeau, op. cit., p. 503
- [27] <https://fr.scribd.com/document/213361885/Les-roles-et-les-objectifs-des-previsions-des-ventes>
- [28] Predictive Analytics World, « *What is Predictive Analytics* »
- [29] Predictive Analytics World, « *Is this a "data mining" conference?* »
- [30] Charles Nyce, « *predictive Analytics White paper* »
- [31] <https://www.axopen.com/modele-predictif/>
- [32] https://fr.wikipedia.org/wiki/Analyse_pr%C3%A9dictive
- [33] <https://www.axopen.com/modele-predictif/>
- [34] Stéphane Tufféry, *Data Mining et statistique décisionnelle*, Paris, éditions Technip, 2010, 705 p.
- [35] Robert Nisbet, John Elder et Gary Miner, *Handbook of Statistical Analysis & Data Mining Applications*, Academic Press, 2009, 823 p.
- [36] <http://cedric.cnam.fr/vertigo/Cours/ml2/coursArbresDecision.html>
- [37] Chaabane LAMICHE, «FUSION ET FOUILLE DE DONNEES GUIDEES PAR LES CONNAISSANCES : APPLICATION A L'ANALYSE D'IMAGE», Mémoire de diplôme de DOCTORAT, UNIVERSITE MOHAMED KHIDER – BISKRA faculté des sciences exactes et des sciences de la nature et de la vie, Année : 2013.

- [38] https://fr.wikipedia.org/wiki/Raisonnement_par_cas
- [39] https://fr.wikipedia.org/wiki/Raisonnement_par_cas
- [40] Béatrice FUCHS, « Représentation des connaissances pour le raisonnement à partir de cas », Mémoire de diplôme de DOCTORAT, Université Jean Monnet de Saint-Etienne, Année : 1997.
- [41] Notes des étudiants du M2IA 2016-2017, reprises et complétées par Amélie Cordier
- [42] S. Tufféry, Data mining et statistique décisionnelle, l'intelligence dans les bases de données, Groupe bancaire Français, Universités de Rennes 1 et paris- Dauphine, 2005.
- [43] <https://www.futura-sciences.com/tech/definitions/internet-java-485/>
- [44] <https://fr.wikipedia.org/wiki/JavaFX>
- [45] https://openclassrooms.com/courses/concevez-votre-site-weeb-avec-php-etmysql/introduction-a_php
- [47] <https://openclassrooms.com/courses/apprenez-a-creeer-votre-site-web-avec-html5-et-css/comment-fait-on-pour-creeer-des-sites-web>
- [48] Nabila BOUSBIA, « Analyse des traces de navigation des apprenants dans un EIAH », année : 2011.
- [49] J. Pernin, A. Lejeune, « Dispositifs d'apprentissage instrumentés par les technologies : vers une ingénierie centrée sur les scénarios », Année : 2004.
- [50] https://fr.wikipedia.org/wiki/Analyse_pr%C3%A9dictive
- [51] https://fr.wikipedia.org/wiki/Analyse_pr%C3%A9dictive
- [52] [https://fr.wikipedia.org/wiki/R%C3%A9gression_\(statistiques\)](https://fr.wikipedia.org/wiki/R%C3%A9gression_(statistiques))
- [53] Charles Manski, « *Regression* », *Journal of Economic Literature*, vol. 29, n° 1, mars 1991, p. 34-50
- [54] Gareth James, Daniela Witten, Trevor Hastie et Robert Tibshirani, *An Introduction to Statistical Learning*, Springer Verlag, coll. « Springer Texts in Statistics », 2013
- [55] https://fr.wikipedia.org/wiki/R%C3%A9gression_lin%C3%A9aire
- [56] https://www.ibm.com/support/knowledgecenter/fr/SSLVMB_24.0.0/spss/regression/idh_nlre.html
- [57] https://fr.wikipedia.org/wiki/R%C3%A9gression_quantile
- [58] Journal Français d'Ophtalmologie Volume 36, n° 8 pages 710-715 (octobre 2013).
- [59] https://fr.wikipedia.org/wiki/Machine_%C3%A0_vecteurs_de_support#cite_note-1
- [60] Bernhard Schölkopf, Alexander J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization and Beyond*, 2002, MIT Press.

[61] https://fr.wikipedia.org/wiki/Analyse_discriminante

[62] <https://fr.wikipedia.org/wiki/WampServer>

[63] http://dictionnaire.sensagent.leparisien.fr/Mozilla%20Firefox/fr-fr/#cite_note-5

[64] <https://www.mozilla.org/en-US/MPL/>

[65] <https://www.generation-nt.com/telechargement-firefox-telecharger-navigateur-web-actualite-948671.html>

[66] https://fr.wikipedia.org/wiki/Windows_10

[67] Anis TRABELSI « Régression de poisson », Romain RESPRIGET

[78] <http://www.sp-mc.com/les-inconvenients-de-la-regression-logistique/>

[79] <http://webapps.fundp.ac.be/umdb/biostats/?q=book/export/html/263>

[70] Loïc BARRAULT, « Machines à Vecteurs Support Séparateurs à Vaste Marge SVM »

Liste des figures

| | |
|--|--|
| Figure 01 : Le triangle pédagogique de J.Houssage | |
| Figure 02 : Une topologie de la formation à distance | |
| Figure 03 : Les éléments de E-Learning | |
| Figure 04 : l'interaction des apprenants avec l'environnement d'apprentissage | |
| Figure 05 : utilisation de modèle prédictif | |
| Figure 06 : représentation d'un arbre de décision | |
| Figure 07 : Noeud d'un réseau de neurone | |
| Figure 08 : hyperplan séparateur..... | |
| Figure 09 : les principales étapes dans le processus d'un système de raisonnement par cas.. | |
| Figure 10 : Contexte Global du système Trails_prédictor | |
| Figure 11 : L'Architecture du système Trails_prédictor | |
| Figure 12 : Module de « Traitement Log» | |
| Figure 13 : Informations du système (PC) | |
| Figure 15 : Page d'accueil de l'application | |
| Figure 16 : Formulaire d'authentification | |
| Figure 17 : l'ouverture du fichier log | |
| Figure 18 : Traitement du fichier log | |
| Figure 19 : Générer les modelés | |
| Figure 20 : prétraitement du fichier log | |
| Figure 21 : Prédiction des parcours d'apprentissage | |
| Figure 22 : Evaluation de la réussite | |

Liste des tableaux

| |
|--|
| Table 01 : Algorithme d'apprentissage par arbre de décision |
| Table 02 : Algorithme qui extrait le modèle de chaque cluster |
| Table 03 : Algorithme de la prédiction des parcours d'apprentissage |
| Table 04 : Algorithme de l'évaluation de la réussite |
| Table 06 : Caractéristiques du système d'exploitation |
| Table 06 : Caractéristiques du système d'exploitation |
| Table 07 : Caractéristiques du système d'exploitation |
| Table 08 : les outils de développement |