Rapport de Stage de fin d'études

Contribution à la Mise en place d'une application de gestion des CVs (CVstore)

Du 01/03/2016 au 31/07/2016

Réalisée par : Farah ARAOUKI

Ingénieure MACS - Mathématiques appliquées et calcul scientifique

Maitre de stage :

Mr Mustapha FONSAU

Tuteur universitaire :

Mr Ahmed KEBAIR

Septembre 2016

Dédicaces

C'est avec un grand plaisir et une telle gratitude que je dédie ce travail : A ma chère famille qui représente mon soutien principal et ma source d'inspiration dans cette vie.

C'est grâce à leur aide, leur encouragement, et leur prière que je réalise aujourd'hui ce mémoire et ce travail. Les mots que je dédie pour eux sont, sans doute, insuffisants pour les remercier de tous ce qu'ils m'ont offerts jusqu'à aujourd'hui. Je vous souhaite tout le bonheur du monde.

A toute ma famille, mes chers amis, mes professeurs et tous ceux qui me sont chers. Que ce travail vous rend fiers de moi.

A.FARAH

Remerciements

Je tiens avant tout à remercier l'école d'ingénieurs sup 'Galilée ainsi que not de m'avoir permis d'effectuer ce stage de 5 mois en entreprise où j'ai pu apprendre auprès d'experts.

Je remercie également mes responsables d'études Monsieur Lafitte Olivier et Monsieur Audusse Emmanuel qui m'ont mis dans les meilleures conditions afin de réaliser mon stage.

Parallèlement, je remercie mon tuteur Monsieur Kebair Ahmed pour sa disponibilité et ses conseils afin de faire de ce stage une véritable plus - value. Mes vifs remerciements accompagnés de toute ma gratitude s'adressent également à Mustapha FONSAU, mon encadrant qui n'a cessé de me conseiller, de me transmettre son savoir. Son attention ainsi que sa disponibilité auprès de moi m'ont permis de développer et d'acquérir de nouvelles compétences.

Je tiens à remercier particulièrement toute l'équipe de Next Challenge, et en premier Mohamed Elkharroubi, qui a su m'accueillir et me mettre dans les meilleures conditions pour que je réalise mon stageEnfin, j'exprime mon grande reconnaissance à tous mes enseignants pour la formation de qualité qu'ils m'ont prodiguée tout au long de mon cursus universitaire au sein de l'Ecole sup Galilée.

Sommaire

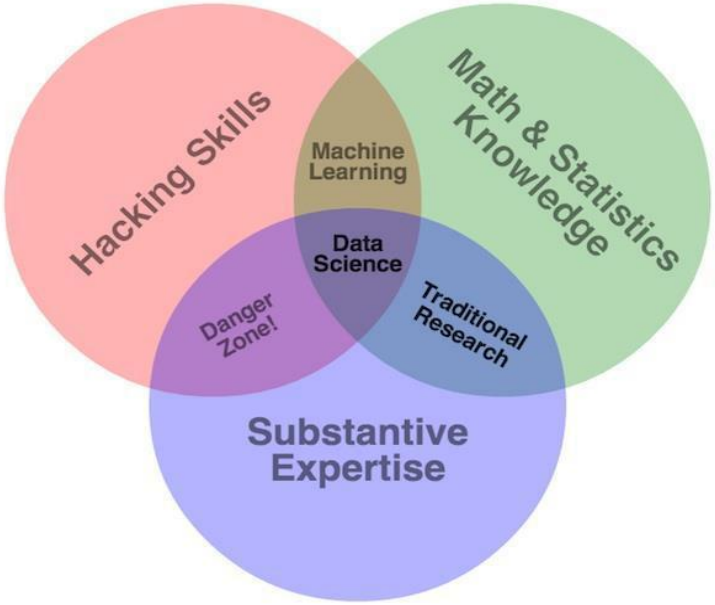
Liste des figures

Liste des tableaux

Introduction Générale

Les bases de données sont le pilier critique des entreprises. Toutes les données liées aux consommateurs et aux fournisseurs sont sensibles et confidentielles. Or la donnée n'est pas statique et évoluent au fil du temps, d'où la nécessité de trouver la solution en fonction de l'état actuelle.

La **science des données** (en anglais ***data science***) est une nouvelle discipline qui allie les mathématiques et l'informatique afin de produire un résultat visuel commode pour aboutir à une conclusion fiable. Ce domaine ne cesse de se développer que ce soit dans le monde universitaire ou au sein des entreprises publics ou privée .



Dans le cadre de ma formation d'ingénieur en mathématiques appliquées et calcul scientifique et dans le but de compléter ma formation par un stage de fin d'études, l'opportunité m'a été offerte d'effectuer un stage de 5 mois, qui s'est déroulé du 1 er mars jusqu'au 31 juillet, au sein de la société NEXT CHALLENGE.

Résumé

NEXT-CHALLENGE est une société spécialisée dans le conseil, le recrutement, et la formation dans les systèmes d'information, adoptant des approches technologiques et managériales novatrices, œuvrant ainsi pour l'émergence de nouvelles formes de consultance, ne stratégie qui lui permit d'avoir une croissance annuelle à deux chiffres.

Mon stage s'est déroulé dans les locaux de NEXT-CHALLENGE à Nanterre, durant la période de mon stage j'ai commencé par une mise en situation et une compréhension du besoin, après une analyse du métier et une structuration du modèle de données je suis passée à une interprétation et analyse des données en faisant une étude comparative des modèles et des algorithmes disponibles ce qui m'as permis de mieux connaitre le métier de

Data science et les enjeux de l'intelligence artificielle (Machine Learning)

Abstract

NEXT-CHALLENGE is a company specialized in consulting, recruitment, and training in information systems, adopting innovative technological and managerial approaches, working well for the emergence of new forms of consultancy, a strategy that allowed him have an annual double-digit growth.

My internship took place in the premises of NEXT CHALLENGE-in Nanterre, in the period of my internship I started with a simulation and an understanding of the need, after an analysis of business and structuring of the data model I was passed to an interpretation and analysis by making a comparative study of models and algorithms available which allowed me to better know the craft of data science and issues of artificial intelligence (machine Learning)

CHAPITRE 1 : INTRODUCTION ET MISE EN SITUATION

I. INTRODUCTION ET MISE EN SITUATION

1. Introduction

Dans ce chapitre, nous traitons le contexte général du projet. Pour y parvenir, nous présentons, tout d'abord, l'organisme d'accueil "NEXT-CHALLENGE". Ensuite, nous expliquons le projet« CVSTORE » en le mettant dans son contexte et en exposant la problématique à résoudre ainsi que notre contribution dans la solution proposée. Enfin, nous décrivons la méthodologie de développement adoptée tout au long de notre travail

2. Présentation de l'entreprise d'accueil

L'entreprise dans laquelle j'ai effectué mon stage se nomme « NEXT-CHALLENGE ».

C'est une start-up française située à Nanterre dans les Hauts-de-Seine (92 000) qui a été créé le 04 Avril 2011 par Mustapha FONSAU qui est également mon encadrant durant mes 5 mois de stage. C'est une société à responsabilité limitée (SARL). Cette société a son siège à Nanterre, où j'ai effectué mon stage, mais est également implantée dans quelques autres villes dans la zone EMEA, comme à Lille (France), à Casablanca (Maroc), à Nouakchott (Mauritanie), à Tunis (Tunisie), à Lomé (Togo) et à Istanbul (Turquie). Le capital de la société est de 10 000€ et en 2015 tandis qu'aujourd'hui son chiffre d'affaires est de 2 800 000€. NEXT-CHALLENGE Nanterre compte aujourd'hui 12 consultants, entre séniors et expert qui œuvrent dans les nouvelles technologies du SI, et fait appel à des dizaines d'indépendants spécialisés et 2 salariés qui sont à la fois commerciaux et formateurs (Mustapha et Aziz, voir organigramme Figure 2). Ce qu'il faut savoir c'est que la société s'agrandit de jour en jour car elle recrute des consultants sur les différents continents. Par ailleurs, la société compte des salariés qui travaillent au siège. Les activités de la société sont réparties comme suit :

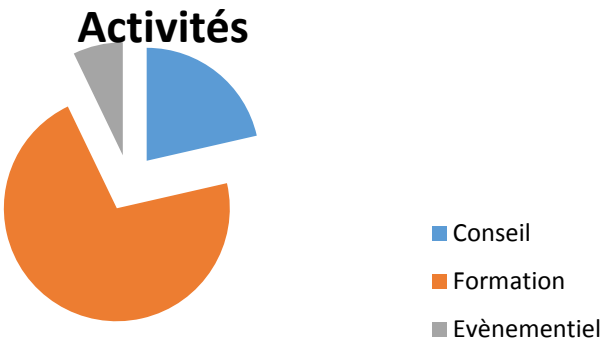


FIGURE 1: ACTIVITES DE NEXT-CHALLENGE

L'équipe de NEXT-CHALLENGE, ayant acquis une expérience longue de 10 ans dans le big data, met à la disposition son savoir-faire dans le domaine du conseil, mangement et la formation pour aider ses clients à concevoir et développer une propre application internet en toute rigueur. NEXT-CHALLENGE intervient notamment chez des clients qui sont des TPE / PME-PME dans le cadre de projets et objectifs prédéfinis initialement. Il est important pour l'entreprise que ses clients ne subissent des débordements de budget ou des erreurs de chiffrage qu'ils doivent absorber. C'est pourquoi NEXT-CHALLENGE recrute des consultants professionnels avec une approche entrepreneuriale.

A présent je vais vous présenter l'organigramme de l'entreprise de Nanterre pour avoir un aperçu

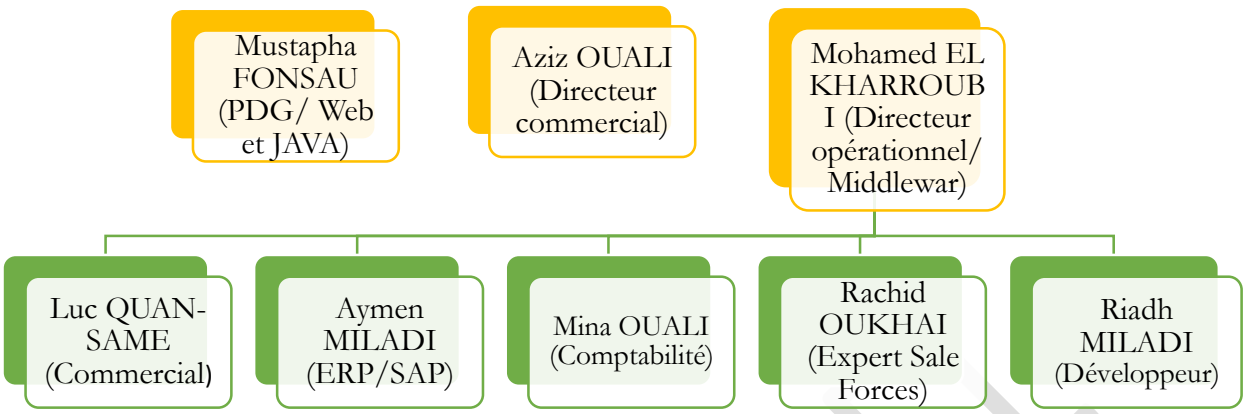


FIGURE 2: L'ORGANIGRAMME DE L'ENTREPRISE

Actuellement, l'entreprise a une stratégie qu'elle tente de développer sur plusieurs axes. Tout d'abord, dans la délégation de formateur c'est-à-dire qu'elle essaye au maximum de trouver et de fournir des formateurs pour ses clients. Ensuite, dans la régie informatique c'est-à-dire que l'entreprise prend en compte les dépenses réels du prestataire qu'elle engage. Autrement dit, cela désigne le placement de salariés chez un client, pour une durée qui peut aller de la journée à plusieurs années, afin de réaliser le travail souhaité par le client. De plus, l'entreprise se spécialise en audit et conseil en stratégie informatique et enfin en prestation informatique. Quant à son positionnement, on peut affirmer qu'actuellement NEXT-CHALLENGE est un acteur majeur dans le milieu du monde de l'informatique.

FIGURE 3: LOGO DE LA SOCIETE « NEXT-CHALLENGE»

3. Contexte

Les entreprises gèrent une masse très grande de données importantes qui ne sont pas gérées de la même façon, elles ne sont pas nécessaires au même moment et ne sont pas attendues sous la même forme.

Cette masse de données est utilisée comme source d'information pour les fonctionnalités opérationnelles et décisionnelles d'une entreprise illustrées par la figure 4 Ces deux volets sont les plus importants à gérer dans n'importe quelle entreprise.

Le volet opérationnel porte sur l'informatisation des différentes opérations quotidiennes qui sont appelées en d'autres termes les processus de l'entreprise. Les informations sont donc au service de ces nombreux processus qui sont généralement faits manuellement d'une façon fatigante. De plus elle génère plusieurs fautes humaines et cause une perte du temps considérable et ne permet pas de générer une trace qui sera utile pour responsabiliser les différents acteurs impliqués.

Quant au volet décisionnel, il est responsable de donner une vue plus globale sur le fonctionnement des processus en offrant des métriques et des informations précises et pertinentes. Cette vue globale aide les décideurs et les acteurs concernés à faire plus facilement des diagnostics et à prendre des décisions très importantes plus facilement et en se basant sur des métriques consistantes et détaillées.

En se basant sur ces deux volets, l'importance de l'informatique dans une entreprise n'est plus à démontrer. Il sera donc judicieux de mettre la force de l'informatique au service de l'entreprise sur les deux volets déjà décrits. Ce qui permet non seulement de prendre de bonnes décisions au bon moment, mais également de gagner du temps en automatisant les différents processus manuels.

Et pour réussir l'informatisation, il convient de choisir de bon matériel et logiciel adéquat puis d'assurer une veille technologique régulière.

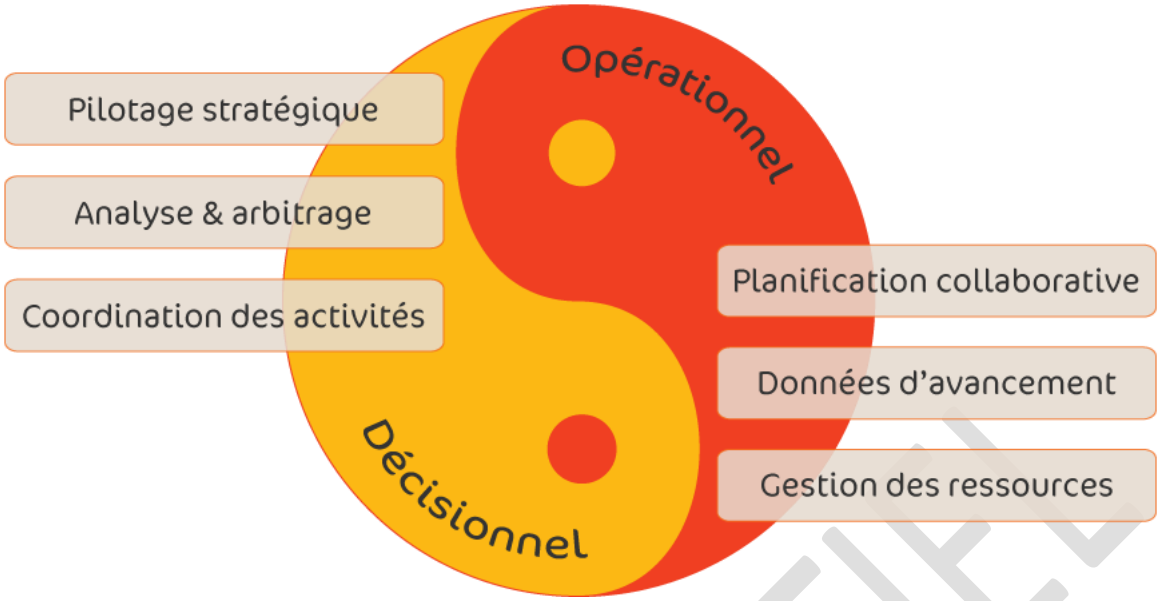


FIGURE 4: LES DEUX PRINCIPAUX VOLETS D'UNE ENTREPRISE

4. Problématique

« NEXT-CHALLENGE» est une startup qui vise, sur le court terme, à doubler son effectif. Ce plan stratégique rend le processus de recrutement et qualification des CV's, l'un de ses processus les plus prioritaires. Cette mission met en évidence tout ce processus : de la sélection initiale de l'employé (formateur ou consultant), son affectation au poste adéquat ou sa participation à la formation chez les clients et à la gestion continue de cette ressource humaine. Il sera judicieux alors de bien gérer ce processus en s'intéressant bien précisément à valider les postes jugés importants, bien filtrer les compétences et trouver rapidement les plus adéquates aux postes proposés afin de répondre rapidement et efficacement aux différents besoins de « NEXT-CHALLENGE ».

La problématique posée par le projet « Réalisation d'une CV thèque pour l'automatisation du processus de recrutement et la classification sémantique des Cvs » réside dans l'absence d'outil au sein de NEXT-CHALLENGE permettant : d'un côté, la centralisation des données, le partage d'informations et plus particulièrement le suivi et l'orchestration de la procédure de recrutement qui est un processus relativement long. D'un autre coté la facilitation de la

recherche complexe des CVs (CURRICULUM VITAE) pour identifier automatiquement les candidatures les plus importantes et pour qu'on puisse les évaluer.

En effet, les processus d'embauche sont ponctués par plusieurs étapes-clé telles que :

- La réception de la demande du Manager (Responsable)
- La prise de contact avec le supérieur hiérarchique pour définir la fiche de poste.
- La diffusion de l'offre d'emploi.
- Le tri des candidatures.
- La recherche des candidatures les plus pertinentes pour chaque poste.
- La transmission des meilleurs dossiers au responsable opérationnel.
- La pré-qualification Téléphonique avec les candidats sélectionnés.
- Les entretiens d'embauche.
- La notification des candidats par l'état de leur demande.
- La convocation des candidats présélectionnés et la prise de décision.
- La variété de structure de CVs reçus qui rend leur extraction et classification très difficile.

A cause de l'importance de ce processus de recrutement/placement des consultants, on doit identifier avec plus de détails les différents problèmes de ce dernier pour pouvoir bien les résoudre au cours de notre projet. Le processus de recrutement souffre alors de :

- La complexité de données reçues : cette complexité est due à la masse de données et à leurs structures diverses qui rendent leur manipulation manuelle fastidieuse et provoque des pertes et des erreurs.
- Ces données reçues quotidiennement sont non seulement d'une grande quantité mais aussi la vitesse de leur réception est très rapide et difficile à gérer.
- Connaitre la disponibilité des consultants, nous sommes obligées de les contacter un par un, cette opération est difficile à gérer.
- La mobilité des consultants représente aussi une information importante à savoir mais qui n'existe pas dans les CVs des candidatures.

Ces raisons ont poussé NEXT-CHALLENGE à repenser tous les processus métier entrant en œuvre, et définir ainsi une solution innovante découlant des différentes étapes des processus

métier, la solution devait aussi implémenter des outils intelligents rendant l'interaction entre les processus très fluide car reposant sur des algorithmes à forte valeurs ajoutés. Ci-dessous les processus métier majeurs du système ainsi que les acteurs qui y sont impliqués:

Processus de pré-sélection :

- Présélectionner les CVs pertinents.
- Ajouter tous les CVs dans notre base
- Automatiser la recherche des CVs les plus adéquats.
- Automatiser la classification des CVs suivant leurs domaines.

Processus de suivi de candidature :

- Automatiser ce processus afin qu'il soit plus rapide, agile et facile à gérer.
- Centraliser les données qui concernent ce processus pour qu'elles soient homogènes, uniques et donc facilement accessibles. Ces données devraient aider à donner une vue globale de ce processus pour prendre les bonnes décisions.

Processus de corrélation entre les ressources et les opportunités business de candidature :

- Identifier les ressources humaines susceptibles de correspondre aux opportunités de placement chez les clients NEXT-CHALLENGE en prenant en considération un nombre important de facteurs.

Présentation de la mission

Ainsi la mission qui m'a été confié au cours de ce stage est la mise en place et l'application des méthodes et algorithmes statistiques et mathématiques dans le cadre de l'évaluation des candidats de Next Challenge et l'amélioration de la recherche et l'analyse prédictive pour notre moteur de recherche et d'indexation géo localisé.

En effet, à l'aide des algorithmes d'intelligence artificielle notre objectif est de classifier les CVs et que en premier temps en attribuant une note à chaque CV et à passer à une comparaison des compétences et d'autres critères comme la géolocalisation le tarif etc.. Afin de proposer aux clients le meilleur choix en réduisant le non réponse à la sollicitation

5. Etude de l'existant : Analyse et critiques

Suite à une recherche approfondie sur les produits qui existent dans le marché, on a pu sélectionner les exemples des plateformes dédiées à la mise en relation professionnelle le plus connus et qui présentent des bonnes références et définissent un modèle à suivre et améliorer durant notre projet.

En effet, la vocation originaire de Linkedin et Viadeo était de mettre en ligne son CV et de créer des connexions avec son réseau professionnel.

Ces exemples sur lesquels j'ai travaillé présentent un résultat d'un grand travail collaboratif afin de servir des grandes structures c'est pour cette raison qu'on a pris du temps pour bien se documenter sur ces projets.

a. Présentation de quelques solutions existantes :

LinkedIn :

C'est un réseau professionnel qui s'appuie sur l'import de nos contacts mail pour proposer de nouvelles relations, il se serre des renseignements saisis dans le profil pour proposer de nouvelles personnes ou des postes qui peuvent nous convenir. LinkedIn aujourd'hui est une méthode efficace pour accroitre son réseau professionnelle et même trouver un emploi adéquat.

Viadéo :

C'est une plateforme presque équivalente à LinkedIn. Elle permet à son utilisateur d'engager des discussions avec des professionnelles et partager des nouvelles ou de publier des articles dans son domaine de compétence afin de se faire connaitre. En créant un profil sur ce réseau, gratuitement, on saisit notre CV mais avec une formule plus structurée afin d'harmoniser le visuel sur tous les profils.

Monster :

Le groupe américain Monster , fort de sa présence dans 40 pays, présente un acteur majeur dans le monde du recrutement en ligne . Dans ce Joboard, on trouve des offres d'emploi très variées proposées par différents types d'entreprises.

Pour faire une recherche d'offre d'emploi, il suffit de les classer soit par métier et/ou par type de contrat et/ou par secteur d'activité. Le résultat se présente sous la forme d'un tableau Facilement exploitable et lisible.

b. Présentation des critères d'évaluation

- **Disponibilité** : Sur Cv-Store il est possible de connaitre la disponibilité du formateur via le calendrier que ce dernier partage sur la plateforme.
- **Evaluation** : Lors d'une formation, le client peut évaluer le formateur (ces ressentis) et y attribuer une note. Cela permettra par la suite aux nouveaux clients de voir les différentes notes qui ont été attribué précédemment au formateur. Ce critère est important car il va permettre au client de connaître par avance les qualités du formateur.
- **Mobilités** : CV STORE renseigne sur la mobilité de ses formateurs et consultants. C'est-à-dire que le client peut savoir si son formateur peut se déplacer, à quel endroit, quand, etc...
- **Géolocalisation** : Cv Store nous permet de géo localiser le formateur et cela présente un avantage pour le client pour se faire une idée sur les formateurs autour de lui.
- **Mots-clés** : Avec CV STORE, il est possible d'ajouter des mots-clés ce qui permet de trouver le formateur ou le consultant le plus adéquat à sa demande.

c. Critique et comparaison

A travers ce tableau comparatif, nous allons pouvoir comparer les différents critères ci-dessus afin d'analyser au mieux les fonctionnalités de notre nouvelle application CV Store.

Critères	Viadeo	Monster	Linkedin
Disponibilités	-	-	-
Evaluation	-	-	-
CV	√	√	√
Formateurs	√	√	√
Intuitif	-	-	-
Mobilités	-	-	-
Géolocalisation	-	√	-
Mots-clés	√	√	-
Prix	-	-	-

TABLEAU 1: ETUDE COMPARATIVE.

Si on prend en compte les critères d'évaluation déjà présentés dans la section précèdante, je constate que CV Store fourni les mêmes fonctionnalités que Linkedin ou Monster, par exemple en ce qui concerne la consultation de CV, la recherche par mots clés, la recherche de formateurs ou encore la géolocalisation.

En revanche, CV Store se distingue par de nouvelles fonctionnalités qui ne sont pas disponibles via les concurrents. En effet, CV Store permet de consulter la disponibilité à un moment précis du profil ou encore prendre en compte des critères d'évaluation et de notation. Le prix est également accessible directement via CV Store ainsi que la mobilité du profil.

Surtout, CV Store permet aussi de proposer des profils de façon intuitive, c'est à dire en prenant en compte l'ensemble des critères automatiquement (géolocalisation, évaluation, disponibilité et prix) afin de proposer immédiatement lors de la recherche, le profil le plus adapté. C'est la force principale de CV Store afin de se démarquer des solutions existantes.

6. Solution proposée

Notre solution consiste à développer une application web légère basée sur les workflows qui traite le processus de suivi des candidatures en s'intéressant à ces deux principales phases :

- **La phase de présélection** : Automatisation de la classification et le filtrage de CVs grâce au processus déjà décrit du TextMining en se basant sur les algorithmes de la classification supervisés offerts par le Machine Learning afin de dégager le domaine adéquat de chaque CV. En d'autres termes, au niveau de cette phase il faut :
  - Extraire les données nécessaires des différents Cvs pour les consulter facilement et rapidement.
  - Donner la possibilité d'une recherche générale pour filtrer les CVs.
  - Donner la possibilité d'une recherche ciblée suivant des critères prédéfinis.
  - Classifier les CVs selon leur domaine, afin d'offrir une recherche plus guidée et plus ciblée.
  - Afficher des tableaux de bord significatifs qui peuvent donner une vue globale sur les différents CVs reçus par domaine.

- **La phase de suivi de candidatures:** Automatisation des étapes de recrutement et de la prise de décision :
  - Extraire les données des candidatures à partir de différents cvs ajoutés pour faciliter leur traitement.
  - Centraliser les différentes informations extraites dans une source unique afin de les bien traiter et les utiliser pour avoir une vue globale de l'entreprise.
  - Charger les différentes informations d'une façon simple et intuitive.
  - Automatiser les processus manuels afin de garantir une coordination fluide, rapide et organisée entre les différents acteurs ce qui va aider à automatiser les notifications, les mails envoyés aux candidats et la planification des entretiens et des tests.

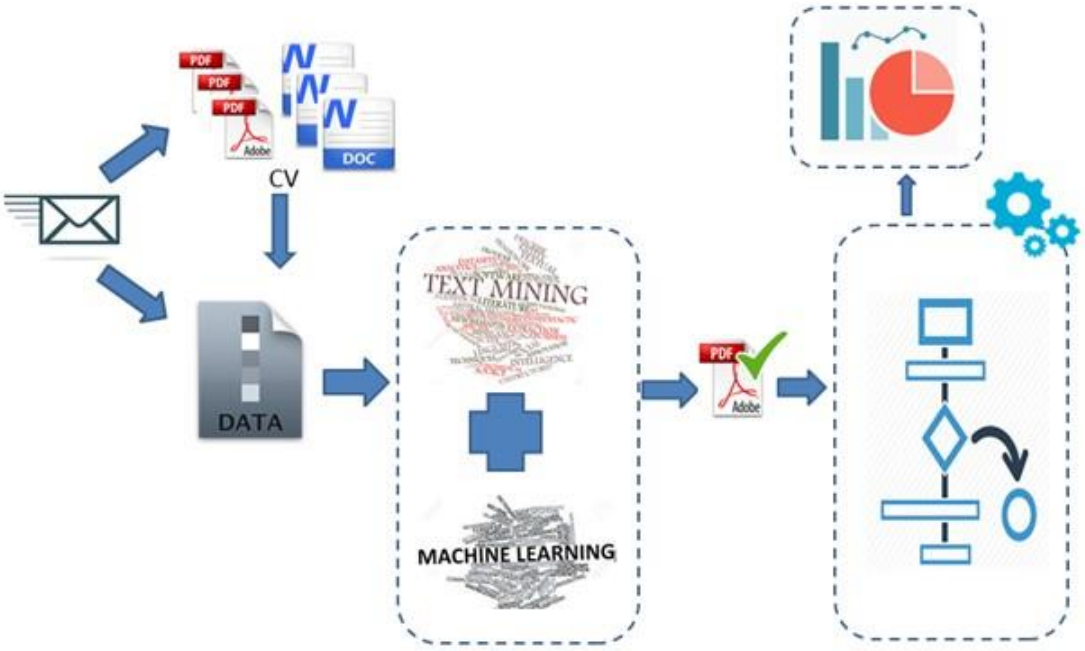


FIGURE 5: VUE GLOBALE DE LA SOLUTION PROPOSEE

7. Gestion du projet (Méthode Agile)

Pour le bon déroulement du projet nous avons besoin d'une méthodologie.

En effet, une méthode agile est une démarche fréquentative. Elle prend en compte le besoin du client pour fournir une solution de me haute qualité répondant aux demandes. L'organisation et l'adoption de la méthode Agile nous a permis d'avoir des résultats très proches des buts préalablement fixés.

Notre choix s'est focalisé sur la méthodologie Scrum,cela nous a permis de prendre en compte l'évolution du besoin du client afin d'améliorer notre stratégie et nous adaper au plus afin de gagner du temps.

La méthodologie : Scrum

i. Présentation de la méthodologie Scrum

Scrum est un processus agile qui peut nour permettre d'atteindre rapidement les objectifs. Une succession de Sprint (une fréquence de 2 à 4 semaines) rythme le développement avec le processus agile. Les activités à élaborer durant ces Sprints sont fixées au préalable en prenant en compte l'urgence des faits et la disposition des membres de l'équipe.

À la fin de chaque sprint, tous les acteurs de SCRUM se mettent au courant de l'état actuel du produit après l'écoulement du temps prévu et prennent la décision de la livraison ou de l'amélioration des réalisations effectuées. La figure 6 nous explique le principe de la méthode Scrum décrit auparavant.

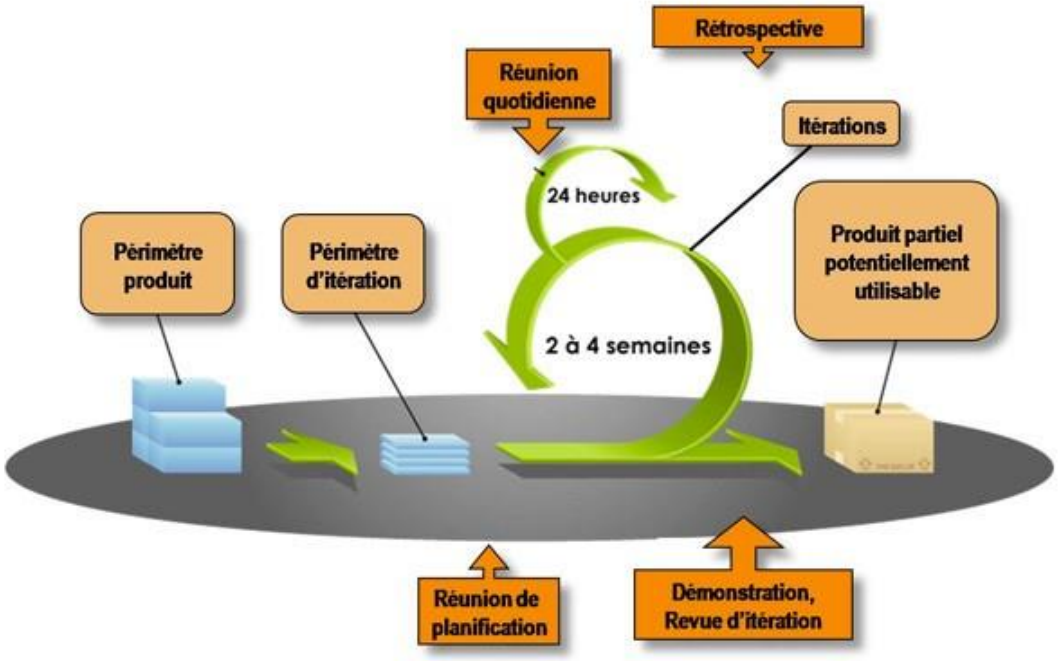


FIGURE 6: PRINCIPE DE LA METHODOLOGIE SCRUM

ii. Les acteurs principaux de Scrum

- Product owner (Représentant des clients et des utilisateurs).
- Scrum Master (Protecteur de l'équipe).
- Scrum Team (L'équipe de développement).
- Stakeholders (Les utilisateurs finaux).

Product owner :Il représente le client et prends les décisions importantes, il définit les besoins et donne son avis sur les réalisations.

Scrum Master : c'est le responsable pricipale face au cliennt. Il doit réussir à maintenir la stabilité au sein de l'équipe et veille à ce que les valeurs de la méthode soient respectées.

Scrum Team :Ce sont les développeurs e l'équipe. Ils sont autonomes et ont le libre choix de la stratégie. A la fin de chaque Sprint, ils doivent présenter un livrable.

- Stakeholders : Ce sont les utilisateurs finaux du produitLa diminution des risques de déviations (Le projet est entièrement développé et testé pour des courtes itérations avec validation du client à la fin du chaque sprint)

8. Conclusion

Dans ce chapitre, nous avons présenté tout d'abord la société d'accueil "Next-Challenge France", ensuite nous avons spécifié les détails de l'application en mettant le projet dans son contexte et nous avons introduit les résultats de l'étude comparative et des critiques pour faciliter la compréhension du projet, et nous avons analysé la situation existante afin de dégager les problèmes à résoudre au cours de ce projet. Les problèmes cités sont présentés avec plus de détails sous forme de besoins fonctionnels et non fonctionnels au niveau du chapitre suivant.

CHAPITRE 2 :

SPECIFICATION ET ANALYSE DU BESOIN

II. SPECIFICATION ET ANALYSE DU BESOIN

1. Introduction

Après avoir mis le projet dans son cadre théorique, nous passons dans cette partie à détailler sa spécification.

La phase d'analyse et spécification des besoins est une étape clé dans la vie du projet. Elle assure la compréhension du besoin en étalant les différentes fonctionnalités que doit remplir le produit. Pour ce faire, dans le présent chapitre nous allons présenter les acteurs de notre application, l'analyse des besoins en détaillant les besoins fonctionnels et non fonctionnels. Ensuite nous allons présenter les différents cas d'utilisation de l'application. Enfin nous exposons quelques scénarios.

2. Identification des acteurs

En s'appuyant sur les besoins exprimés précédemment, nous identifions quatre acteurs principaux qui devront être en interaction continue avec notre application Ces acteurs sont :

- **Next-challenge RH** :

Il doit disposer d'un compte pour pouvoir se connecter à l'application. Il peut accéder et bénéficier des fonctionnalités qui lui ont été offertes par l'application à savoir le suivi du processus de demande d'un consultant, la gestion d'encaissement et filtrage des CVs etc.

- **Next-challenge négociateur:**

C'est un processus manuel qui permet au Next-challenge négociateur de fixer tarif journalier d'un consultant selon la formation.

- **Le Client:**

c'est un acteur (soit client (entreprise) soit formateur) qui a comme rôle de s'authentifier pour consulter son profil dans le cas d'un formateur ou bien de réserver et évaluer dans les cas d'une entreprise a la recherche d'un formateur

- **Le Consultant :**

Introduit durant le cas d'utilisation pour la négociation du tarif journalier de la formation.

3. Spécification des besoins

La spécification des besoins nous permet de définir les besoins applicables à un système pour fournir les services dont les utilisateurs ont besoin.

Dans cette section, nous allons préciser les différents besoins fonctionnels et non fonctionnels de l'application afin de modéliser à travers des diagrammes UML.

a. Les besoins fonctionnels

Les besoins fonctionnels décrivent les fonctionnalités de l'application. Ils sont les besoins spécifiant un comportement d'entrée/sortie du système.

BF1- Gestion des demandes d'emploi :

Elle doit assurer l'automatisation du processus de demande d'emploi.

BF2- Présélection des CVs adéquats par une recherche et filtrage des CVs :

L'application doit permettre aux différents acteurs de :

- Filtrer les CVs suivant les compétences.
- Filtrer les CVs suivant leur université.
- Filtrer les CVs suivant les langues.
- Permettre au décideur de faire des recherches des termes ou d'expressions bien déterminées dans les CVs
- visualiser les termes les plus récurrents dans l'ensemble des CVs par domaine.
- Chercher les profils les plus adéquats aux postes demandés.
- Consulter les CVs par domaine.
- Classer les universités dont les candidats ont envoyé le plus de demandes.

BF3- Gestion du suivi d'affectation a une mission ou une formation:

L'application doit assurer le suivi du processus de recrutement pour les candidats présélectionnés.

b. Besoins non fonctionnels

Les besoins non fonctionnels représentent les contraintes implicites auquel le système doit répondre. Parmi lesquelles nous citons :

- **Évolutivité et extensibilité :**

Proposer une architecture permettant l'ajout des fonctionnalités futures aisément.

- **Fiabilité :**

L'application doit fonctionner en se basant sur des sources de données cohérentes.

- **Temps de réponse raisonnable :**

Le temps de réponse à la fonctionnalité demandée par l'utilisateur doit être raisonnable.

- **Les tableaux de bord :**

Les tableaux de bord ne doivent pas être surchargés et il faut bien les présenter afin de mettre en valeur les informations pertinentes

- **Sécurité :**

Respecter les politiques internes de gestion de la sécurité et de la confidentialité.

4. Analyse des besoins

a. Analyse Métier :

L'Analyse Métier doit permettre de définir minutieusement les besoins afin d'apporter une solution adéquate.

L'Analyse Métier doit être capable de faire face aux changements afin de pouvoir adapter les stratégies et les décisions tout en essayant d'intégrer une touche d'innovation pour se différencier des autres.

Dans le cadre de notre projet, l'analyse métier détecte le besoin et cherche la meilleure structure pour mieux attaquer le problème de classification des CVs puisque ce dernier ne cesse d'évoluer durant la vie professionnelle d'une personne, ainsi que les offres d'emploi.

Nous préciserons en premier temps les critères d'évaluation d'un CV, ensuite nous proposerons une méthode qui répond au besoin en améliorant la recherche sur Cv store

b. Structuration du modèle de données :

Dans le but de définir la structure des données et extraire le maximum d'informations des CV, nous avons mis en place un formulaire à remplir par des candidats de Next Challenge dont le but de croiser les informations qui concernent les candidats et les exigences des missions.

i. Que recouvre le concept de Master Data Management ?

Les données de chaque entreprise sont structurées dans une base de donnée sou au sein d'un logiciel métier .La mise à jour des données est donc effectuée par plusieurs équipements dans des entités différentes. Cette méthode peut provoquer une incohérence entre les différents changements. C'est bien là l'objectif de la méthode de "Gestion des données de base".

ii. Comment fonctionne cette méthode ?

Elle réside dans la collecte des données de l'entreprise. Cela permet d'avoir un référentiel qui décrit la dépendance et les liens entre les différentes données de l'entreprise afin de garder l'harmonisation de ces datas et de ne pas créer de doublons. Ce but est atteint en passant par des étapes de contrôles dynamiques.

Libellé	Descriptif du libellé
Coordonnées	**Indiquer le contact**
Nom*	Le nom du consultant / formateur
Prénom *	Le prénom du consultant
Age *	L'âge du consultant
Adresse *	Ville, code postal, Pays
Situation familiale *	Célibataire, marié, avec ou sans enfants.
Contact*	Numéro téléphone, mail
Permis de travail*	préciser le lieu (national, international)
Résumé**	Un résumé de carrière qui a un minimum de 50 caractères et 500 caractères maximum
Domaine d'intervention/compétences	**Liste de domaine d'intervention et compétences**
Compétences techniques*	Les compétences du consultant par thèmes, par niveau maîtrise (NM)
Langues**	Indiquer la langue et NM
Keywords	Par mots clés, et nombre d'occurrence, pertinence
Expériences professionnelles	**Liste des expériences professionnelles**
Durée *	Date de début (mois et année) -date de fin de la mission
Environnement (technique)*	Les outils utilisés, indiquer NM, développement, matériel, logiciels
Domaine *	Dans quel domaine a été effectuée la mission
Entreprise*	Le nom de la société, informations
Secteur d'activité*	Préciser le secteur
Formations / certifications /diplômes	**Liste de formations, certifications & diplômes**
Formations	Formations effectuées (année)
Certifications	Indiquer les certifications obtenues
Diplômes*	Diplômes obtenus (année)
Tarif *	**liste des tarifs (TJM)**
Prix pour formations	Indiquer le TJM vente selon le type de prestation donnée/ compétence
Prix proposé par le consultant	indiquer le TJM selon les compétences
Marge	Marge de gain : 20% et 80€ minimum
Disponibilités :	**Indiquer les disponibilités du consultant**
Calendrier *	partager l'agenda afin de voir les disponibilités du consultant/formateur
Mobilité *	Sa zone de mobilité, la distance acceptée
Coordonnées bancaires*	Joindre une PJ de RIB
Informations sur l'entreprise**	Siret, Siren, Code SIRET Code NAF, Nom de la structure Assujetti ou non à la TVA Attestation de versement à l'URSSAF
Formations données	les formations et coaching

TABLEAU 2: UN APERÇU DU MDM DE CV STORE

En plus des différents paramètres et informations extraites d'un CV d'autres critères complèteront le MDM tels que le poids de la donnée et sa valeur.

5. Conclusion

Dans ce chapitre, nous avons précisé les différents besoins fonctionnels et non fonctionnels à satisfaire par notre application. L'expression des différents besoins procure une vision plus claire du sujet et une compréhension plus profonde des tâches à réaliser. Nous avons également précisé les méthodes de structure du modèle de données L'étape suivante, présentée dans le prochain chapitre, permet d'étudier les algorithmes mathématiques qui répondront à la problématique de classification des CVs.

CHAPITRE 3 :
MODELISATION ET INTERPRETATION

III. MODELISATION ET INTERPRETATION

1. Introduction et définition de besoin

Après avoir exprimé les besoins fonctionnels et non fonctionnels de notre application, nous abordons la partie interprétation et analyse de données ainsi que l'étude de modèle et algorithmes disponibles

A partir d'un MDM, nous avons défini un formulaire qui sera proposé aux consultants /formateurs lors de l'inscription dans notre base de données et qui comportera : des informations personnelles, les compétences, l'expérience professionnelle, son tarif etc. …

1ère fouille de Données_: A partir des différentes types d'informations saisies par le consultant/formateur notre but en premier lieu est d'attribuer une note à tous les consultants/formateurs qui dépendra de plusieurs paramètres : tels que les compétences techniques, l'expérience dans un tel domaine etc. …

Dans un premier temps, nous allons analyser les différentes données (complètes /incomplètes) appartenant à un formateur/consultant nous devrons trouver la bonne formule pour les évaluer

Pour résoudre notre problème, nous avons eu recours aux réseaux bayésiens pour modéliser la problématique en utilisant un modèle graphique, qui définit la structure de données et qui se base sur les probabilités conditionnelles ce qui traduit un critère très important dans notre analyse qui est le critère de dépendance qui signifie que l'évaluation d'un consultant/formateur peut changer si on change n'importe quel paramètre

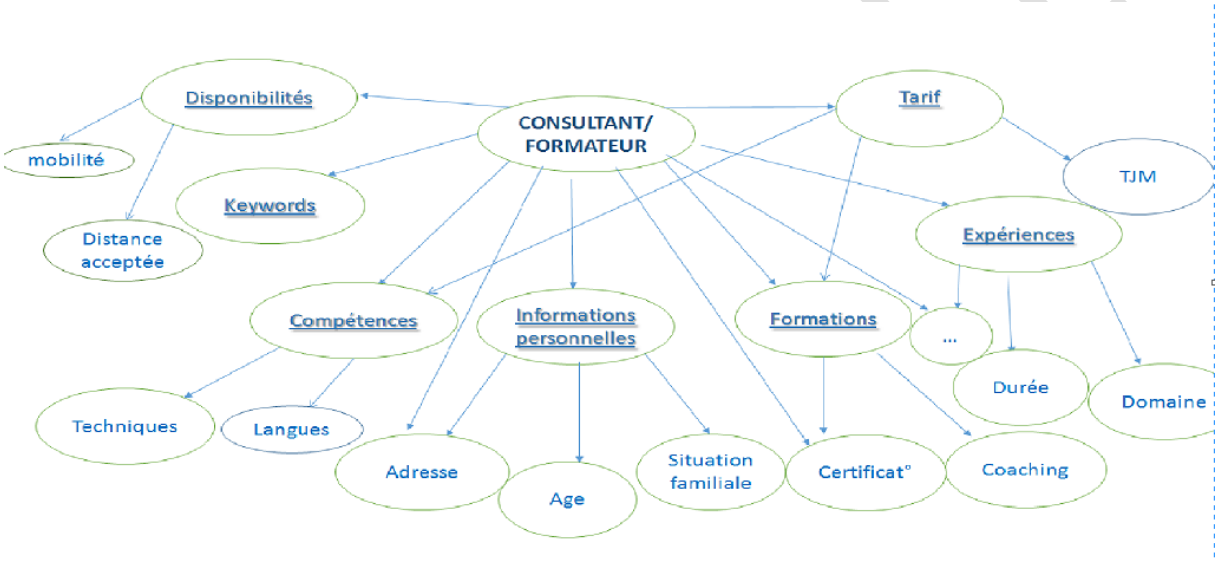
D'où le rôle des graphes dans les modèles probabilistes et statistiques est triple :

1. faciliter la mission des experts,
2. Simplifier le calcul des probabilités jointes,
3. Aide à l'étude de la partie inférence à partir des données

2. ETUDE DES ALGORITHMES DISPONIBLES

a. Définition

Les réseaux bayésiens sont considérés comme des modèles graphiques. Dans le but de faciliter le calcul des probabilités jointes d'un système de variables aléatoires, ce modèle allie deux théories des graphes et des probabilités et donne un moyen capable de faire le calcul nécessaire. Un critère très important est mis en valeur dans cette méthode est celui de la causalité qui se trouve entre les différents acteurs de ce modèle, autrement dit chaque variable est présente une distribution qui traduit la relation causale



Ci-dessous un exemple de réseau bayésien qui pourra modéliser le problème :

FIGURE 7: EXEMPLE DE RESEAU DE MODELISATION

Les étapes de construction de ce modèle sont trois :

- Étape qualitative :

Une des qualités de cette méthode est de donner la possibilité aux experts de construire un modèle qui répond à leurs problématiques sans se soucier des calculs numériques

La figure ci-dessus représente le réseau bayésien modélisant notre problème : le nœud consultant/formateur peut prendre un nombre de valeurs correspondant aux états possibles dans le cas de l'évaluation d'un candidat. L'arc allant du nœud tarif au nœud compétences exprime le fait que le tarif dépend directement des compétences.

- ❖ Étape probabiliste :

L'étape probabiliste est l'étape pendant laquelle se font les calculs probabiliste de notre modèle à l'aide en prenant en considération les différentes distributions de probabilités des variables constituant le modèle.

$$P(n1, n2, \dots, np) = \prod_{i=1}^{p} P(n_i | pa(n_i))$$

Où les n_i sont les variables de notre modèle

- ❖ Étape quantitative

Arrivant à ce stade de résolution de problème, nous procéderons à la détermination des tables de probabilités des variables de notre système. En effet, il faut procéder à la mise en place des table de probabilité qui contiennent les lois de probabilités des variables ainsi que leurs états possibles..

b. Planification du réseau

Dans le but de construire notre réseau, nous allons construire une base d'exemple sur laquelle on va appliquer des algorithmes de structure afin d'obtenir notre structure de données
Dans notre cas, la base d'exemples décrit l'état du réseau au cours du transfert des données d'un nœud à un autre. Cette base d'exemples se présente sous la forme d'une matrice, appelée« data », formée de n lignes (nombre de nœud) et m colonnes (nombre de mesures effectuées).

Pour déterminer cette base d'exemples, nous avons réalisé l'algorithme ci-dessous

---

Algorithme

- filling the cost matrix by data

- 10% of all the arcs are considered defective

- M = Ø

- Choose the size of the data T

- For i =1 to T do

Choose nodes : em = transmitter node

rec = receiver node

Apply Dijkstra algorithm on the cost matrix

nbrn = (passage number of nodes between em and rec) +1,

nbra = number of arcs visited during transmission between em and rec

If transmission between em and rec is directAlors dir = 2

Then dir = 1

If transmission between em and rec shortest path route

So court = 2 et pbarc = 1

Then court = 1 et pbarc = 2

End for

---

c. Apprentissage de la structure de données

La détermination de la structure des données est une phase assez complexe si elle n'est pas déterminée par un expert du domaine, cette structure peut être produite par un algorithme qui part d'une base d'exemple pour définir le squelette de travail en donnant le graphe initial sur lequel s'effectueront les calculs d'inférence. Pour répondre à cette question, nous avons étudié les principaux algorithmes disponibles.

i. Les algorithmes disponibles

Nous allons étudier les 5 méthodes suivantes :

MWST : l'algorithme de l'arbre de recouvrement de poids minimal (minimum weight spanning tree) le but de cette méthode est de remplir la matrice des poids qui correspond au poids attribué à chaque arête potentiel. il suffira après d'appliquer un des algorithme de résoltuion de ce genre de problème comme l'algorithme de Kruskal ou celui de Prim.

PC : recherche de causalité : le principe de cette méthode est déterminé les indépendances conditionnelles entre les paramètres, on met en place un test si l'indépendance existe on supprime l'arc correspondant

K2 : Le but de cet algorithme est d'ordonner les paramètres en attribuant un ordre à chaque variable par exemple un une fois on fixe un ordre il ne sera pas possible de prendre un nœud comme parent de son ascendant .L'algorithme K2 supporte aussi l'optimisation de la structure

GS : l'algorithme de recherche gloutonne : dans cette méthode il s'agit de prendre un graphe initial et de le comparer avec ses voisins, ensuite choisir le bon graphe et passer à une autre itération. Le bon graphe est celui qui a un score maximal.

SEM : il est basé sur le principe de l'algorithme Expectation-Maximisation (EM) et il assure le traitement des bases de d'exemple manquantes par exemple comme dans notre cas nous avons retiré 20% des données pour tester l'algorithme.

Dans ce qui suit, nous allons tester les différents algorithmes permettant de définir la structure des données à partir d'une base d'exemples et de comparer les résultats avec le graphe d'origine.

ii. Comparaison des algorithmes :

Nous allons tout d'abord définir les critères de comparaison :

La distance d'édition : le nombre d'opérations nécessaires pour transformer le graphe obtenu en celui d'origine.

Complexité : définit la complexité de l'algorithme

Score : $BIC(\mathcal{B}, D) = \log P(D|\mathcal{B}, \theta^{M,V}) - \frac{1}{2} Dim\ (\mathcal{B}) \log N$

Où D est notre base d'exemples, $\theta^{M,V}$ est la distribution des paramètres obtenue par *maximum de vraisemblance* pour le réseau $\mathcal{B}$ et où $Dim\ (\mathcal{B})$ est la dimension du réseau bayésien

Stabilité : Dans notre cas nous appelons la stabilité est le nombre de données utilisées pour se rapprocher du réseau original.

Nous avons effectué une première série de test sur les différentes méthodes d'apprentissage de structure de données en prenant en question la distance d'édition et les autres critères.

iii. Résultats et interprétation

Une première série de tests nous a permis d'évaluer la précision de ces méthodes en essayant de retrouver le graphe. Les résultats obtenus montrent qu'il est difficile de retrouver des relations "faibles" entre les variables avec peu d'exemples. L'initialisation aléatoire de la plupart des méthodes peut aussi être remplacée efficacement par une initialisation issue d'une première de recherche utilisant une méthode simple et rapide comme l'algorithme MWST

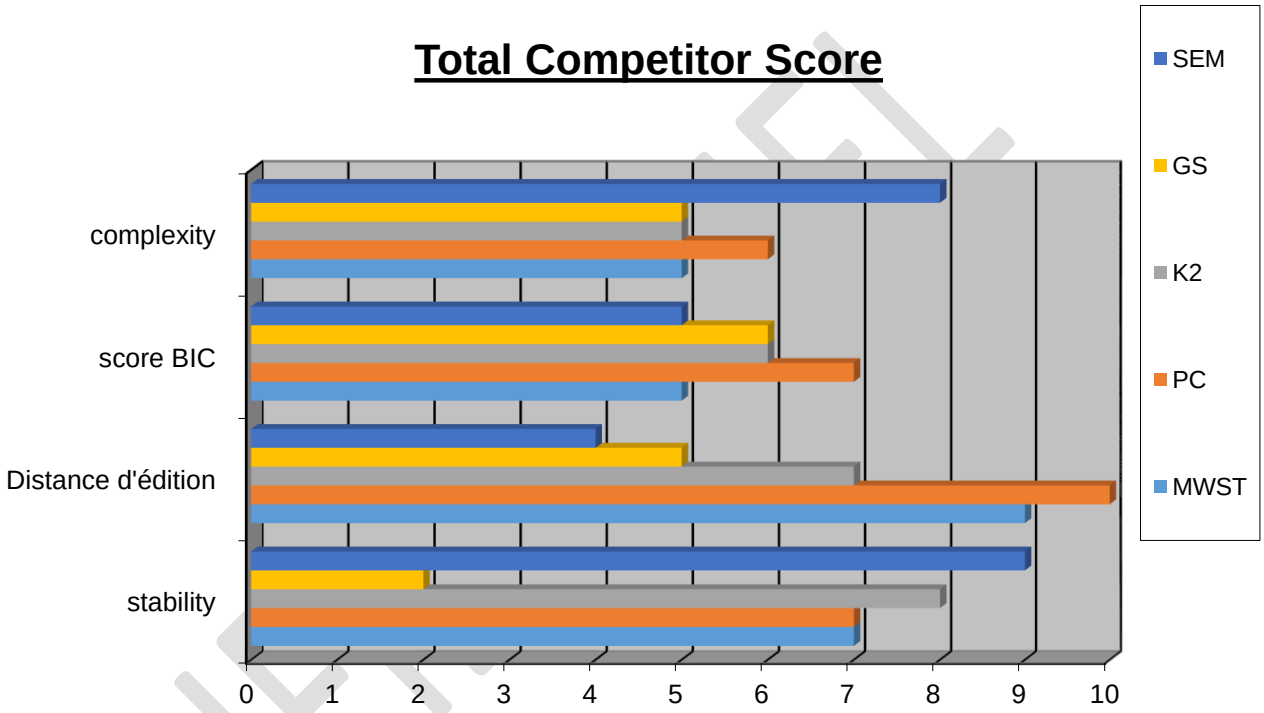


FIGURE8 : SCORE TOTALE CONCURRENT

Afin de tester l'efficacité des mêmes méthodes face à des problèmes de classification.

La figure 8 présente le résultat d'une étude comparative des algorithmes disponibles. Nous remarquons que chaque méthode se caractérise par un critère mais la méthode MWST nous a surpris car malgré sa simplicité elle permet d'aboutir à des résultats généralement obtenu par des algorithmes beaucoup plus complexes. Notons aussi la méthode SEM est utilisé dans le cas des données incomplètes et donnent une bonne approximation.

Voir annexe 1

d. Conclusion

Afin de choisir le bon algorithme pour notre problème nous avons étudié l'application de certains algorithmes des réseaux bayésiens sur un échantillon d'exemples et après interprétation des résultats obtenus, nous allons adopter l'algorithme MWST qui sera utilisé dans l'algorithme de l'arbre de jonction dans la phase qualitative.

3. Inférence bayésienne

Le but des réseaux bayésiens est de résoudre les problèmes de classification et prédiction. Dans notre cas, nous cherchons à interpréter les données observées et de les comparer avec les données initiales d'où l'inférence bayésienne est le calcul de P(B|A) où A est l'ensemble d'observations et B un ensemble de paramètres qui expliquent le problème permettant le diagnostic et la prédiction du système. Pour appliquer l'inférence, nous avons étudié les deux types d'inférences :

a. Inférence exacte

Il s'agit de calculer la distribution de probabilité d'un réseau de variables en partant d'un modèle observé , dans ce cadre nous trouvons la méthode d'arbre de jonction ainsi que l'algorithme cut-set conditionning qui se base des données observé mais qui ne sera pas applicable pour des problème densément connecté

b. Inférence approximative

Les méthodes d'inférences ne sont pas praticables pour de réseaux de taille importante. Ce qui explique l'utilisation des méthodes approximatives dans le cas de grand réseau. Comme les algorithmes basés sur l'échantillonnage aléatoire (Monte Carlo) tel que dont la précision dépend de la taille des données test.

c. Conclusion

Un réseau bayésien est un graphe dirigé acyclique : les nœuds correspondent à des variables aléatoires et chaque nœud a une distribution conditionnelle. C'est un moyen de voir les indépendances entre les différents paramètres constituant un graphe.

Ci-dessous un tableau comparatif des différentes méthodes d'inférence que nous pouvons appliquer dans un réseau bayésien

Méthodes exactes (ou complètes)	Méthodes approchées
Algorithme de l'arbre de jonction : La diffusion des informations pour établir la cohérence. La complexité de l'algorithme JT est exprimé en fonction de la taille de sa plus grande clique, ce type d'inférence peut ne pas être praticable dans le cas des grands réseaux **Algorithme cut-set conditionning :** Le même problème se présente ici qui est le problème de la taille du réseau mais l'avantage de ce cas est que sa complexité en temps linéaire	Ici on considère que le graphe était un arbre : "**loopybelief propagation"** **Markov chain Monte Carlo**: utilise la géométrie du graphe en effectuant un échantillonnage de Gibbs sur les sous clusters. **Inférence variationnelle**: cette méthode est une exploitation de l'algorithme EM (Expectation-Maximization).Il est de plus en plus pratiqué

TABLEAU 3: METHODES D'INFERENCE

CHAPITRE 4 : ALGORITHMES UTILISES ET OUTILS

IV. ALGORITHMES UTILISES ET OUTILS

1. Introduction

Dans cette parie nous allons présenter et détailler le fonctionnement de l'algorithme de l'arbre de jonction que nous avons appliqué à la structure de donnée ainsi que les outils informatiques utilisé dans la partie d'implémentation

2. Algorithme de l'arbre de jonction dit JLO

a. Définition

L'algorithme de l'arbre de jonction est un algorithme d'intelligence artificielle (machine Learning en anglais). Il est exploité dans la théorie des modèles graphiques.

Il s'agit d'un ensemble de clique arrangé de façon que la multiplication des fonctions potentielles soit égale à la probabilité conjointe de l'ensemble des variables. L'algorithme fonctionne comme suit :

– *la phase de construction* : Il s'agit dans cette première phase de partir d'un graphe initial et le but est de créer l'arbre de jonction construite par des cliques (une sélection des nœuds).

Cette étape est très importante pour l'élimination des boucles du graphe, d'un côté et pour rendre le graphe plus puissant dans le sens où nous allons gagner en temps de calcul pendant la phase d'inférence

On décompose le processus en trois étapes :

– la moralisation du graphe

– la triangulation du graphe

– la création d'un arbre couvrant minimal, appelé arbre de jonction.

– *la phase de propagation :*

Cette phase est dédiée pour le calcul probabiliste des différentes variables avant et après passage de message de façon à faire circuler les nouvelles informations dans tout le réseau et mettre à jour toutes les variables. Il s'agit d'envoie et réception des nouvelles informations entre les cliques de l'arbre jusqu'à ce que tout l'arbre soit mise à jour.

La phase de construction

Moralisation

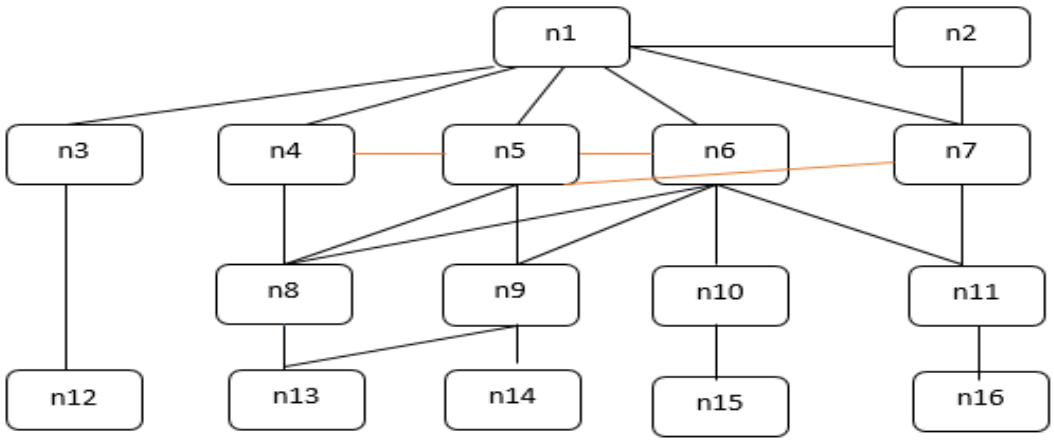
La modification du graphe commence par la moralisation du graphe : il s'agit de combiner les ascendants des variables en les connectant par un arc non-dirigé. A la fin de cette étape, on enlève les directions des arcs rajoutés ce qui nous donne un graphe non-dirigé, G^m est appelé « graphe moralisé »

Triangulation

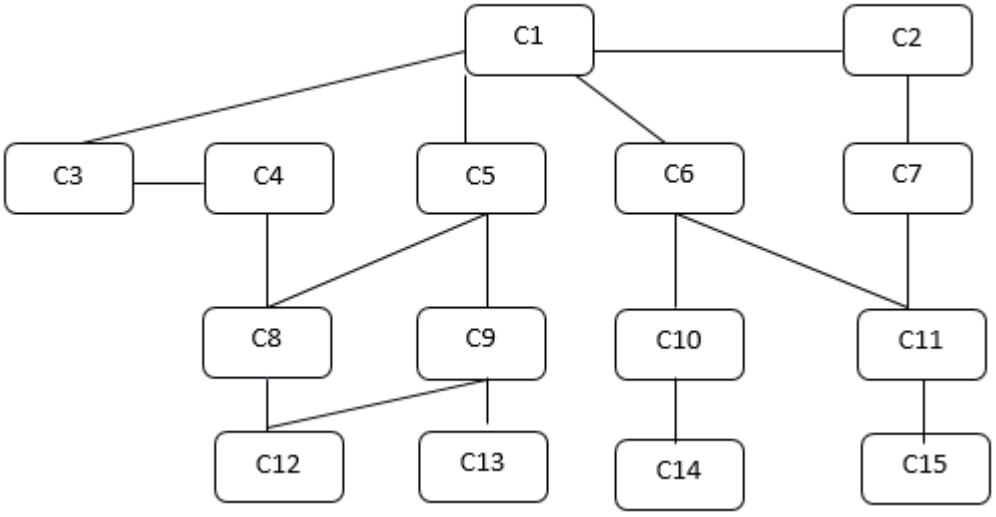
Ensuite, on passe à la triangulation du graphe moral : l'idée ici est de construire des cliques de variables, qui sont des regroupements de clusters du graphe moralisé G^m . Ces clusters seront les nœuds de l'arbre de jonction sur lequel s'effectueront les calculs dans une phase postérieure. D'où l'obtention du graphe dit triangulé G^T .

Arbre de Jonction

Pour finir avec cette partie de création de l'arbre de jonction depuis le graphe initial :



Graphe moralisé



Arbre de jonction

a. Initialisation de l'arbre de jonction

A cette étape du travail, on se base sur un arbre de jonction déjà construit. Ce stade de l'étude se fait une fois uniquement. Les phases que nous allons aborder porteront sur le calcul des probabilités dans les réseaux bayésiens

Cette phase comporte les caractéristiques du graphe G (graphe initial) dans le but est de calculer une spécification numérique adéquate pour l'arbre de jonction.
La distribution jointe de probabilité pour un graphe non-dirigé G^u peut être exprimée sous la forme suivante :

$$P(X_1, \dots, X_n) = \prod_{C \in \mathcal{C}} a_C(x_C)$$

Où $\mathcal{C}$ est l'ensemble des cliques du graphe G^u, x_C est l'ensemble de valeurs attribuées aux variables de la clique C et les fonctions a_C sont des fonctions non-négatives dont les valeurs sont dans l'ensemble des affectations possibles des valeurs des variables de la clique C et donnent une valeur dans l'intervalle $[0, \infty[$

On considère à ce niveau l'intégralité des séparateurs associés à chaque couple de cliques adjacentes dans l'arbre construit précédemment, on attribue à chaque séparateur une fonction de potentiel b_S que nous définissons de façon semblables aux fonctions de potentiel a_C . Or,

par définition, un séparateur est l'intersection de deux cellules voisines, la distribution de probabilités jointe associé au réseau initial peut s'écrire de la façon suivante :

$$P(X_1, \dots, X_n) = \frac{\prod_{C \in \mathcal{C}} P(x_C)}{\prod_{s \in S} P(x_s)}$$

. L'égalité ci-dessus est utile dans le calcul de l'inférence dans les réseaux bayésiens avec l'algorithme de l'arbre de jonction

Le principe est simple, on attribue chaque variableX_i à une clique de l'arbre. Il se peut que certaines cliques restent sans aucune variable attribuée .Une fois, nous avons attribué toutes les X_i , nous définissions le potentiel de la manière suivante

$$a_C(x_C) = \begin{cases} \prod_i P(X_i | pa(X_i)) \quad si \ X_i \quad est \ mis \ dans \ \mathrm{C} \\ 1 \ si \ aucune \ variable \ n'est \ dans \ \mathrm{C} \end{cases}$$

b. La phase de propagation

Le but de cette étape de l'algorithme est de faire circuler les nouvelles informations dans l'arbre en passant d'une clique à une autre voisine dans l'arbre et de rafraichir les informations des cliques voisines ainsi que des séparateurs avec ce nouveau message.

Ce qui est intéressant dans cet algorithme c'est que la distribution $P(X_1, X_2, \dots\dots, X_n)$ associé au graphe triangulé reste bonne après chaque mise à jour des cliques. Si toutes les informations locales ont été transmises, cette phase de l'algorithme prendre une nouvelle forme qui est une la représentation des fonctions de probabilités de l'arbre sachant son état initial et les nouvelles informations présentes

i. Flux d'information entre les cliques :

Dans cette partie nous allons étudier le processus de passage des flux entre les cliques :

$$P(U) = \frac{\prod_{C \in \mathbb{C}} a_C(x_C)}{\prod_{s \in S} b_s(x_s)}$$

En effet, la façon avec laquelle les informations circulent est la suivante.

On définit le *flux* d'une clique C_i à une clique voisine comme suit.

Soit S_k le séparateur de ces deux cliques, alors

$$b^*_{S_K}(x_{S_K}) = \sum_{C_I \setminus S_k} a_{C_i}(x_{C_i})$$

La mise à jour de la clique C_i se fait de la façon suivante :

$$a^*_{C_j}\left(x_{C_j}\right) = a_{C_j}(x_{C_i})\lambda_{S_k}(x_{S_k})$$

Avec

$$\lambda_{S_k}(x_{S_k}) = \frac{b^*_{S_k}(x_{S_k})}{b_{S_k}(x_{S_k})}$$

Le terme $\lambda_{S_k}(x_{S_k})$ est appelé le *facteur de mise à jour*. Un message est une nouvelle information que reçoit la clique C_i , elle est connu par le séparateur S_k qui représente la partie partagée entre C_i et C_j , le but est de transmettre les message reçu par C_i à sa voisine C_j en passant par le séparateur d'où le flux qui le passage des informations depuis C_i vers toutes les autres clique constituant l'arbre

Le phénomène d'ordonnancement est défini comme suit : définir une clique comme racine de l'arbre, et faire passer les messages depuis les feuilles vers cette racine (sachant que n'importe quel nœud peut être considéré comme racine de l'arbre c'est ce qu'on appelle la phase de collection

ii. Entrer une évidence dans le réseau

La question à laquelle il faut répondre maintenant c'est comment faire rentrer une évidence (nouvelle information dans le réseau) ? En effet, à chaque fois qu'un nouveau message se présente, il est introduit dans l'arbre et circule dans tout le réseau

Pendant la phase d'insertion de nouvelles informations, il est possible d'introduire autant d'évidence que le nombre de variables.

Autrement dit : une *évidence* est définie $\alpha : \omega \rightarrow \{0,1\}$ tel que les états des variables ω ne sont pas tous possible. Si un seul état est prouvé possible alors après avoir circulé dans tout le réseau, la variable qui a recueilli l'information aura une probabilité égale à 1

iii. Le cout ce cette phase

La complexité de cette phase de l'algorithme est exprimé en fonction du nombre de clique ainsi qu'on nombre de ses états elle est de l'ordre de $O\left(\sum_{i=1}^{N_c} n_e\,(C_i)\right)$.

Donc si nous cherchons à améliorer cette complexité.il sera judicieux de composer des cliques avec le minimum de variables simple (avec un faible nombre d'état).Néanmoins la résolution d'un tel problème (arbre optimal avec des cliques de taille minimale est un problème NP-difficile

3. Environnement de travail

Nous présentons dans ce qui suit, notre environnement de travail. Au début, nous décrivons l'environnement matériel puis l'environnement logiciel qui a permis à l'aboutissement de la mise en œuvre de l'application pour enfin indiquer les outils et le langage choisi.

a. Environnement matériel

Pour concrétiser notre application, nous avons utilisé un PC ayant les caractéristiques suivantes :

Caractéristiques	Type
Processeur	Intel® Core i5-5005U
Marque	Apple iMac
Fréquence	3.1 GHz
RAM	8 GO
Système	Mac OS X
Disque dur	1 TO
Carte graphique	Intel Iris Pro Graphics 6200

TABLEAU 4: CARACTERISTIQUES DU PC UTILISE

b. Les logiciels mises en œuvre

i. MongoDB

Comme beaucoup de sites dynamiques, celui-ci a besoin d'une base de données.

Dans la plupart des cas, on utilise une base de données relationnelle mais dans le but de mieux manipuler les documents et éviter la duplication des données : avoir deux bases de données

(une pour les utilisateurs et l'autre pour les documents, il nous a été imposé d'utiliser MongoDB.

FIGURE 9: LOGO MONGODB

Ce qui m'a permis de comprendre la logique de manipulation des documents à travers le NoSQL.

ii. Apache Solr

Apache Lucene est un moteur de classification de documents texte qui permet d'effectuer la fouille en langage original grâce à des différentes manipulations automatiques du texte.

Le document indexé est sauvegardé sous plusieurs représentations, et de même pour le texte affecté, et les résultats de l'exploration sont donnés après l'analyse de ces variantes.

Ce moteur élargie le principe Lucene en simplifiant la gouvernance (interface RESTful) et en rapportant des fonctionnalités : filtres de recherche, traitement des résultats, etc.

FIGURE 10: LOGO SOLR

iii. H2O.ai

H2O est un logiciel open source pour le big-data analysis. H2O est utile pour appliquer facilement des algorithmes mathématiques et l'analytique prévisionnel pour résoudre des problèmes commerciaux les plus provocants d'aujourd'hui.

Il combine intelligemment les caractéristiques uniques pas actuellement trouvées dans d'autres plates-formes d'apprentissage automatique.

FIGURE 11: LOGO H2O.AI

iv. Langage R

R est un logiciel libre de manipulation des données et d'étude statistiques en consacrant le langage de développement S. Il s'agit d'un outil GNU basé sur les outils crés dans les laboratoires Bell par John Chambers et ses collaborateurs.

Le logiciel R est utilisé par ses développeurs analogiquement à une expropriation de S, tout en gardant la logique du langage Scheme . C'est un logiciel libre attribué avec la licence GNU GPL et disponible sousGNU/Linux, FreeBSD, NetBSD, OpenBSD, Mac OS X et Windows

FIGURE 12: LOGO LANGAGE R

4. Conclusion

Au cours de ce chapitre nous avons présenté et argumenté nos choix techniques sur lesquels nous nous sommes basés durant la réalisation de notre application. Nous avons, ensuite, présenté les principales interfaces élaborées au cours de ce projet.

CONCLUSION GENERALE

L'automatisation des taches et l'analyse des données, aussi bien internes qu'externes, permet à l'entreprise d'offrir un environnement agile et d'améliorer la collaboration entre ses employés, et orienter les décisions prises.

Nous avons, donc, réalisé une solution permettant, principalement, d'automatiser le suivi des CVS reçues, de classifier un gros volume de CVs et d'analyser leur contenu pour fournir des tableaux de bord contenant différents indicateurs utiles lors de la prise de décision.

Pour développer cette solution, nous avons eu recours à H2O for Big Data pour l'extraction des données des candidatures et des CVs à partir des mails reçus du serveur de messagerie de l'entreprise et leur stockage dans une base de données de type MongoDB. Des workflows sont mis en place à l'aide d'Activité pour automatiser le suivi des profils des consultants/formateurs par une application web basée sur AngularJS, coté client et H2O.ai et solr coté serveur.

Quant aux CVs, ils sont classifiés et analysés à l'aide de Spark et solr et les résultats sont retournés à la base MongoDB.

Étant parvenu à mettre en place la chaine quasi complète de la solution souhaitée, il est, maintenant, possible d'optimiser chaque phase de ce processus et d'y ajouter de nouvelles fonctionnalités notamment dans la partie intelligence artificielle au niveaux de l'arbre de jonction construite dans le cadre de la méthode des réseaux bayésiens et pourquoi pas passer aux réseaux bayésiens dynamiques qui peuvent offrir de nouvelles possibilités d'évaluation des candidats en tenant en comptes d'autres critères d'évaluation.

ANNEXES

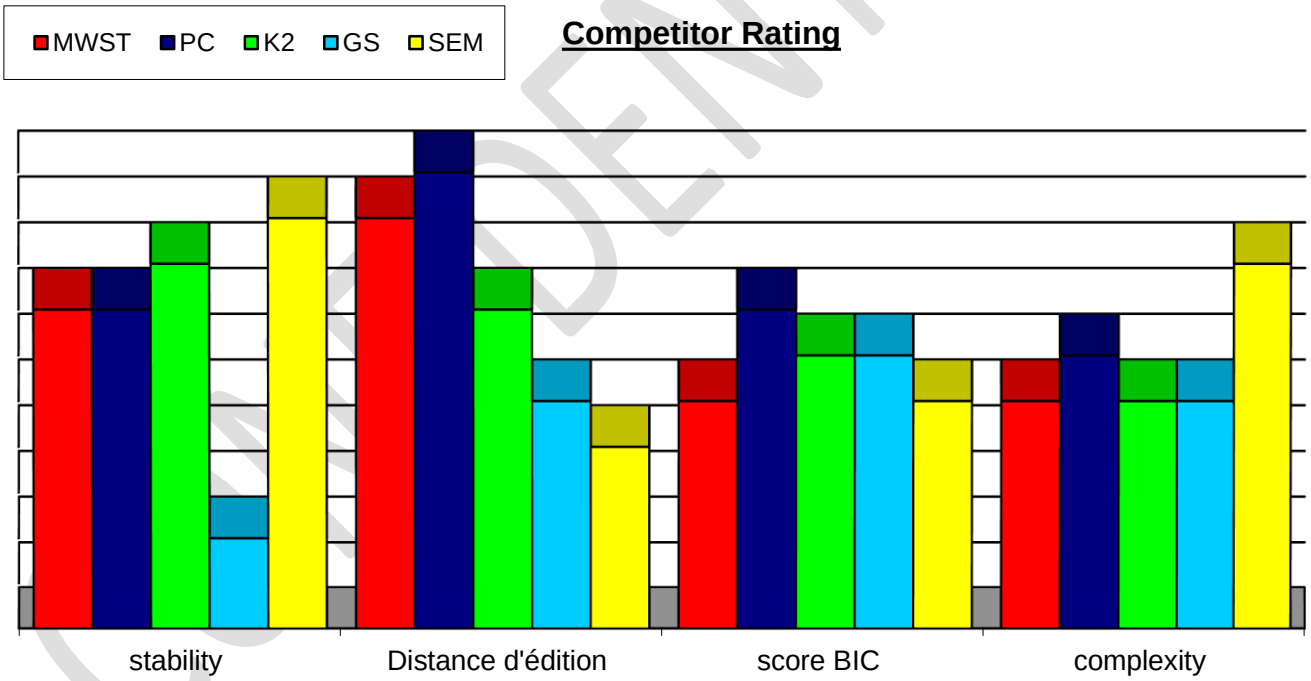


Figure annexe 1 : étude comparative des algorithmes de structure de données

BIBLIOGRAPHIE

[1] Mme. Veronique Messager Rota. Méthodologie Agile. http://www.qualitystreet.fr/2007/11/20/ methodes-agiles-un-belle-definition/.

[2] Mme. Veronique Messager Rota. Les acteurs principaux de Scrum. http//images.slideplayer.fr/3/1291291/slides/slide_45.jpg/,

[3] Mr Kent Beck. Les nouveautés de Scrum. www.agilemanifesto.org/iso/fr/.

[4] Mr Jean-Loup Kars.Quels stacks sont les plus utilisés parmi 500 startups tech française https://www.linkedin.com/pulse/ quels-stacks-sont-les-plus-utilis,.

[5] Mr David Bellot https://tel.archives-ouvertes.fr/tel-00009190/document

http://david.bellot.free.fr/research/Inferences%20dans%20les%20Reseaux%20Bayesiens%20-%20David%20Bellot.pdf

http://www.setit.rnu.tn/last_edition/setit2009/Information%20Processing/182.pdf

[6] Camille Séka Kotchi, Véronique Delcroix et Sylvain Piechowiak http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.1721&rep=rep1&type=pdf

[7] http://www2.ift.ulaval.ca/~lamontagne/ift17587/modules/module5/r%C3%A9seauxBayesiens.pdf

[8] O. Francois, Ph. Leray Evaluation of structure learning algorithms for bayesian networks