



Arbres Généalogiques

Projet Arbres Généalogiques - Semestre 2

**Groupe 18 GMF2
Langevin Maelann
Pichon Alex
Bulard Samuel
Pailhé Thomas
Pierrat Thomas
Ducasse Clément**

Mai 2024

Sommaire

1	Introduction	4
2	Introduction Arbres Généalogiques	5
3	Applications de Généalogie les Plus Connues	6
3.1	Ancestry.com	6
3.2	MyHeritage	6
3.3	FamilySearch	6
4	Analyse des Avantages et Inconvénients	7
4.1	Avantages	7
4.2	Inconvénients	7
5	Cahier des Charges	8
6	Objectifs	10
7	Fonctionnalités Principales	11
7.1	Gestion du Fichier CSV	11
7.2	Consultation de l'Arbre Généalogique	11
7.3	Prédiction des Risques de Diabète	11
8	Interface Utilisateur	12
9	Développement	13
9.1	Environnement de développement	13
9.1.1	Langage de programmation	13
9.1.2	Bibliothèques de Java	13
10	Introduction diabète	14
11	Classification du diabète	15
11.1	Diabète de type 1	15
11.2	Diabète de type 2	15
11.3	Diabète gestationnel	15
12	Dépistage du diabète	16
13	Machine Learning	17
13.1	Apprentissage supervisé	17
13.2	Apprentissage non supervisé	17
13.3	Apprentissage avec renforcement	18
14	Méthodes de Prédiction du Diabète Utilisant le Machine Learning	19
15	Méthode qui sera utilisée K-means	20
15.1	Notations, Mesure de distorsion	20
15.2	Algorithme	20
15.3	Convergence et initialisation	20
15.4	Choix de K	20
16	Règles et contraintes mise en place	21
16.1	Vérification des données	21
16.2	Intégrité des Données	21
16.3	Gestion des Anomalies	21
16.4	Maintenance et Mise à Jour	21
16.5	Contrainte de Performance	21
17	Conclusion	22

1 Introduction

La généalogie, l'étude des liens de parenté et des ascendances, est une pratique ancestrale qui a traversé les siècles. Avec l'avènement de la technologie, cette discipline a connu une évolution remarquable, passant des registres papier aux bases de données numériques et aux applications informatiques spécialisées. Dans ce contexte, notre projet vise à développer une application innovante dédiée à la création et à la gestion d'arbres généalogiques, offrant aux utilisateurs une plateforme intuitive et efficace pour explorer et documenter leur histoire familiale.

L'objectif principal de cette application est de permettre aux utilisateurs de tracer et de gérer les liens familiaux de manière précise et organisée. À travers l'importation et la manipulation de données provenant d'un fichier CSV, notre application permettra de construire un arbre généalogique global, représentant les relations entre les membres de la famille, y compris les parents, les enfants, les frères et sœurs, et d'autres proches.

La première partie du projet se concentre sur la manipulation du fichier CSV contenant les informations relatives aux membres de la famille. Des opérations telles que la vérification de la cohérence des données, le nettoyage du fichier, l'ajout et la suppression de lignes seront mises en œuvre pour garantir l'intégrité et la fiabilité des données.

Ensuite, notre application offre une variété de fonctionnalités pour consulter et explorer l'arbre généalogique global. Les utilisateurs peuvent visualiser différentes perspectives de l'arbre, y compris l'ascendance et la descendance d'une personne donnée, ainsi que les relations familiales telles que les frères et sœurs. De plus, des fonctionnalités avancées sont proposées, telles que la prédiction des risques de survenue d'un diabète en utilisant des méthodes de regroupement telles que l'algorithme K-means.

Enfin, l'interface utilisateur conviviale de notre application offre une expérience agréable et intuitive pour naviguer et manipuler l'arbre généalogique global. Les utilisateurs peuvent facilement visualiser et mettre à jour les informations relatives à leur famille, facilitant ainsi la préservation et le partage de leur héritage familial.

Dans ce rapport, nous présenterons une étude approfondie des applications de généalogie existantes, suivie d'une description détaillée de notre solution conceptuelle et de son implémentation en Java. Nous fournirons également des tests unitaires couvrant les différentes fonctionnalités de l'application, ainsi qu'un jeu de données de test au format CSV. Enfin, nous mettrons à disposition le code source de notre application sur GitHub, avec une documentation complète pour faciliter la compréhension et la maintenance du projet.

2 Introduction Arbres Généalogiques

Un arbre généalogique est une représentation graphique de la généalogie ascendante ou descendante d'un individu, dit de *cujus* (celui sur lequel porte la généalogie). Par abus de langage, cette représentation structurée des liens familiaux entre les personnes est souvent appelée arbre à l'image de l'arbre végétal mais il existe également d'autres représentations par exemple circulaire ou semi-circulaire. Par extension, ce type de schéma peut être utilisé pour les taxons de tout type d'être vivant, même si on parle plutôt de généalogie pour les humains et de pedigree pour les animaux.

Les applications de généalogie offrent divers avantages et inconvénients en fonction de leurs fonctionnalités, de leur convivialité et de leur capacité à gérer efficacement les données généalogiques.

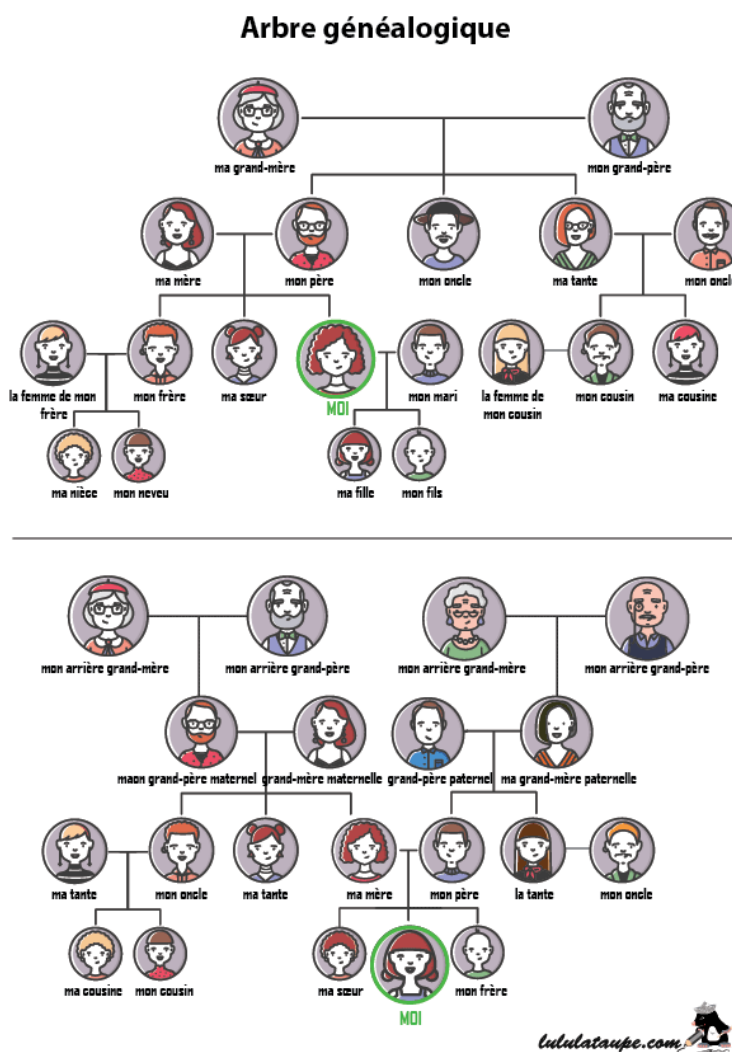


Figure 1: Arbre généalogique

3 Applications de Généalogie les Plus Connues

Parmi les applications de généalogie les plus connues, voici les trois qui se distinguent particulièrement :

3.1 Ancestry.com

Ancestry est souvent considéré comme la principale référence dans le domaine de la généalogie en ligne. Il propose une immense base de données d'archives historiques, ainsi que des outils avancés pour construire et explorer des arbres généalogiques interactifs.

3.2 MyHeritage

MyHeritage est une plateforme populaire offrant une combinaison de recherche généalogique, d'outils de construction d'arbres familiaux et d'analyses ADN pour aider les utilisateurs à découvrir leur héritage familial et à trouver des parents biologiques.

3.3 FamilySearch

FamilySearch, géré par l'Église de Jésus-Christ des Saints des Derniers Jours, est une ressource gratuite et accessible à tous. Il offre une vaste collection de documents historiques et des outils de recherche généalogique, en mettant l'accent sur la collaboration et le partage d'informations entre les utilisateurs.

4 Analyse des Avantages et Inconvénients

Voici une analyse des avantages et des inconvénients des applications de généalogie les plus connues :

4.1 Avantages

- Facilité d'utilisation : Les meilleures applications de généalogie offrent une interface conviviale qui permet aux utilisateurs de créer et de gérer facilement des arbres généalogiques.
- Collaboration : Certaines applications permettent la collaboration entre plusieurs utilisateurs, ce qui est utile pour les familles élargies travaillant ensemble sur leur histoire familiale.
- Intégration de médias : Les applications modernes permettent souvent d'ajouter des médias tels que des photos, des vidéos et des documents scannés, ce qui enrichit l'expérience généalogique en donnant vie aux ancêtres.
- Fonctionnalités avancées : Certaines applications offrent des fonctionnalités avancées telles que des outils de recherche intégrés pour découvrir de nouveaux ancêtres, des options de cartographie pour suivre les déplacements des ancêtres, et des analyses génétiques pour explorer les origines ethniques.

4.2 Inconvénients

- Coût : De nombreuses applications de généalogie nécessitent un abonnement payant pour accéder à toutes leurs fonctionnalités, ce qui peut être prohibitif pour certains utilisateurs.
- Dépendance aux données en ligne : Les applications qui dépendent fortement des bases de données en ligne peuvent poser des problèmes de confidentialité et de sécurité des données, en plus de nécessiter une connexion Internet constante.
- Limitations de personnalisation : Certaines applications peuvent être limitées dans leur capacité à gérer des structures familiales complexes ou à prendre en charge des données généalogiques provenant de sources non traditionnelles.
- Courbe d'apprentissage : Les applications les plus complètes peuvent avoir une courbe d'apprentissage raide, ce qui peut décourager les utilisateurs novices.

5 Cahier des Charges

Un cahier des charges pour un projet d'arbre généalogique décrit les exigences, les objectifs, les ressources nécessaires, les étapes de réalisation et les critères de succès du projet. Voici comment créer un cahier des charges détaillé pour un arbre généalogique :

1. Introduction :

- Objectif du projet : Décrire l'objectif principal de l'arbre généalogique. Par exemple, "Créer un arbre généalogique complet et précis de la famille Dupont sur plusieurs générations."
- Contexte : Fournir un contexte sur la raison d'être du projet, comme l'intérêt familial, historique ou culturel.

2. Description du projet :

- Portée du projet : Définir la portée en termes de générations à couvrir, branches familiales spécifiques, etc.
- Limites : Établir les limites du projet, par exemple, ne pas inclure les membres de la famille éloignés ou les familles par alliance.

3. Objectifs et résultats attendus :

- Objectifs spécifiques :
 - Collecter et vérifier les données familiales jusqu'à X générations.
 - Créer un arbre généalogique visuel et interactif.
- Résultats attendus :
 - Un arbre généalogique imprimable.
 - Une version numérique interactive.

4. Exigences fonctionnelles :

- Collecte des données :
 - Noms complets, dates de naissance, de mariage, de décès.
 - Photos, métiers, anecdotes familiales.
- Sources de données :
 - Actes d'état civil, registres paroissiaux, entretiens familiaux, documents historiques.
- Outils et logiciels :
 - Utilisation de logiciels de généalogie comme Family Tree Builder, Gramps, ou des plateformes en ligne comme Ancestry, MyHeritage.
- Format de présentation :
 - Arbre généalogique imprimable.
 - Version numérique (site web, application).

5. Exigences techniques :

- Logiciels requis : Détails des logiciels à utiliser pour la collecte, l'organisation et la présentation des données.
- Matériel requis : Scanner pour numériser les documents, ordinateur, imprimante de haute qualité.
- Format des fichiers : CSV, GEDCOM pour l'import/export de données généalogiques.

6. Ressources :

- Équipe de projet :
 - Généalogistes amateurs ou professionnels.
 - Développeurs pour la version numérique.
 - Graphistes pour la version imprimable.
- Budget : Estimation des coûts pour les logiciels, abonnements à des bases de données, impressions, etc.

- Calendrier : Chronologie détaillée des étapes du projet avec des dates de début et de fin pour chaque tâche.

7. Étapes du projet :

- Phase de planification : Définir les objectifs et le périmètre.
- Phase de recherche : Collecte et vérification des données.
- Phase de développement :
 - Création de l'arbre généalogique initial.
 - Développement de la version numérique interactive.
- Phase de validation : Vérification des données et des représentations graphiques.
- Phase de publication : Impression et mise en ligne de l'arbre généalogique.

8. Critères de succès :

- Complétude : Toutes les branches familiales spécifiées sont incluses.
- Précision : Les informations sont vérifiées et exactes.
- Accessibilité : L'arbre est facilement accessible aux membres de la famille.
- Satisfaction des parties prenantes : Feedback positif de la famille.

9. Risques et contraintes

- Risques :
 - Difficulté à obtenir certaines informations.
 - Contraintes de temps et de budget.
- Stratégies d'atténuation :
 - Plan de secours pour les données manquantes.
 - Suivi régulier du budget et du calendrier.

6 Objectifs

- Concevoir une application conviviale permettant aux utilisateurs de créer, visualiser et manipuler des arbres généalogiques.
- Permettre l'importation de données à partir d'un fichier CSV pour construire un arbre généalogique global.
- Assurer la cohérence et l'intégrité des données après chaque mise à jour du fichier CSV.
- Proposer des fonctionnalités avancées telles que la visualisation de différentes parties de l'arbre, la prédiction des risques de diabète et la gestion des clusters de membres de la famille.

7 Fonctionnalités Principales

7.1 Gestion du Fichier CSV

- Vérification de Cohérence : L'application doit vérifier la cohérence des données du fichier CSV lors de son importation, en s'assurant que les relations familiales sont valides et qu'il n'y a pas de doublons.
- Nettoyage du Fichier : Une fonctionnalité de nettoyage automatique doit être disponible pour supprimer les données invalides ou redondantes du fichier CSV.
- Ajout et Suppression de Lignes : Les utilisateurs doivent pouvoir ajouter de nouvelles personnes au fichier CSV ainsi que supprimer des personnes existantes.

7.2 Consultation de l'Arbre Généalogique

- Visualisation de l'Arbre Généalogique Global : Les utilisateurs doivent pouvoir visualiser l'arbre généalogique complet à partir des données du fichier CSV.
- Consultation par Famille : Une fonctionnalité permettant de visualiser une partie spécifique de l'arbre généalogique relative à une famille donnée doit être disponible.
- Descendance et Ascendance : Les utilisateurs doivent pouvoir visualiser la descendance et l'ascendance d'une personne donnée de manière claire et intuitive.
- Relations Familiales : La possibilité de visualiser les frères, sœurs et autres proches d'une personne donnée doit être offerte, en proposant une liste exhaustive des liens de parenté possibles.
- Personnes sans Ascendant : Une fonctionnalité pour afficher la liste des personnes sans ascendant doit être incluse.

7.3 Prédiction des Risques de Diabète

- Enrichissement de l'Arbre Généalogique : Les données pertinentes telles que les antécédents de diabète et les facteurs de risque associés doivent être ajoutées à l'arbre généalogique global.
- Regroupement des Membres de la Famille : Utilisation de l'algorithme K-means pour diviser les membres de la famille en clusters homogènes en fonction de leurs caractéristiques de risque de diabète.

8 Interface Utilisateur

- Interface Conviviale : L'application doit offrir une interface utilisateur conviviale et intuitive, facilitant la navigation et l'utilisation des fonctionnalités.
- Manipulation de l'Arbre : Les utilisateurs doivent pouvoir effectuer des manipulations telles que l'ajout, la suppression et la modification des personnes dans l'arbre généalogique.
- Recherche et Filtrage : Des fonctionnalités de recherche et de filtrage doivent être fournies pour permettre aux utilisateurs de trouver rapidement des personnes spécifiques dans l'arbre.

9 Développement

9.1 Environnement de développement

- **Eclipse IDE:** Eclipse est un environnement de développement intégré (IDE) largement utilisé pour le développement Java. Il offre une gamme complète d'outils pour la création d'applications Java, y compris la gestion de projet, le débogage, le refactoring, et l'intégration avec d'autres outils de développement. Eclipse est apprécié pour sa flexibilité et sa grande communauté de développeurs.

9.1.1 Langage de programmation

- **Java:** Java est un langage de programmation polyvalent et robuste, largement utilisé pour le développement d'applications logicielles. Il offre une syntaxe claire, une gestion automatique de la mémoire et une forte typisation, ce qui en fait un choix populaire pour les projets de toutes tailles et complexités.

9.1.2 Bibliothèques de Java

- **JavaFX:** JavaFX est une bibliothèque graphique moderne pour la création d'interfaces utilisateur riches dans les applications Java. Elle offre des fonctionnalités avancées pour la conception d'interfaces utilisateur interactives et attrayantes.
- **Apache Commons CSV:** Apache Commons CSV est une bibliothèque Java pour la manipulation de fichiers CSV. Elle offre des fonctionnalités avancées pour lire, écrire et manipuler des données au format CSV dans les applications Java.

10 Introduction diabète

Le diabète sucré (ou diabète par abus de langage) est une maladie liée à une défaillance des mécanismes biologiques de régulation de la glycémie (concentration de glucose dans le sang) menant à une hyperglycémie chronique.

Cette maladie se manifeste par des symptômes propres au diabète (syndrome polyuropolydipsique) et par des lésions d'organes tels la rétine, les reins ou les artères coronaires, dues à la toxicité de l'acétone produite lors d'une dégradation rapide des graisses arrivant en cas de défaillance de l'insuline menant à une hyperglycémie majeure.

11 Classification du diabète

La classification du diabète selon l'Organisation mondiale de la santé (OMS) est globalement classée en trois catégories :

- **Diabète de type 1**
- **Diabète de type 2**
- **Diabète gestationnel**

11.1 Diabète de type 1

Regroupe le diabète principalement attribuable à la destruction des cellules bêta du pancréas, qui s'accompagne d'une carence en insuline susceptible d'évoluer vers une acidocétose diabétique. Cette forme de diabète comprend les cas attribuables à un processus auto-immun et les cas dont la cause de la destruction des cellules bêta est inconnue.

11.2 Diabète de type 2

Les personnes atteintes de diabète de type 2 sécrètent de l'insuline, mais cette hormone régule avec moins d'efficacité le taux de sucre dans leur sang. Ce taux, appelé glycémie, reste anormalement élevé après un repas, ce qui est la définition du diabète. Petit à petit, le pancréas s'épuise à sécréter des quantités croissantes d'insuline. Également appelé diabète gras, ou diabète non insulino-dépendant, le diabète de type 2 touche surtout les personnes en surpoids ou obèses, sédentaires, le plus souvent après 45 ans.

11.3 Diabète gestationnel

Le diabète gestationnel, aussi appelé diabète de grossesse, est une situation d'intolérance au glucose qui apparaît chez des femmes enceintes qui auparavant n'avaient aucun antécédent diabétique. Chez certaines, la maladie, qui se caractérise par une intolérance au glucose entraînant une glycémie supérieure à 1,26 g/l de sang, va persister au-delà de la grossesse. Le diabète gestationnel concernerait environ une grossesse sur 20. La grossesse est en elle-même diabétogène, mais la résistance au glucose ne se révèle pathologique que dans une minorité des cas.

12 Dépistage du diabète

Toute personne ayant des membres de sa famille atteints de diabète de type 2 doit se faire dépister régulièrement, car un risque héréditaire existe. Si l'un des deux parents est diabétique de type 2, le risque héréditaire est de 40%. Si les deux parents sont atteints, le risque monte à 70%. Pour le diabète de type 1, le risque de transmission aux enfants est de 6% si le père est diabétique, 2 ou 3% si la mère l'est, et 30% si les deux parents sont atteints de diabète. Les personnes en surpoids ou souffrant de troubles de la glycémie doivent également se plier au dépistage. Il en va de même pour les femmes ayant développé du diabète pendant leur grossesse (diabète gestationnel) ou ayant mis au monde un bébé de faible poids. Le dépistage est également recommandé aux personnes de plus de 65 ans.



Figure 2: Test de diabète

13 Machine Learning

Le Machine Learning est une technique de programmation informatique qui permet aux ordinateurs d'apprendre par eux-mêmes sans programmation explicite. Il utilise des probabilités statistiques pour améliorer leur mode d'apprentissage et leurs connaissances de façon autonome.

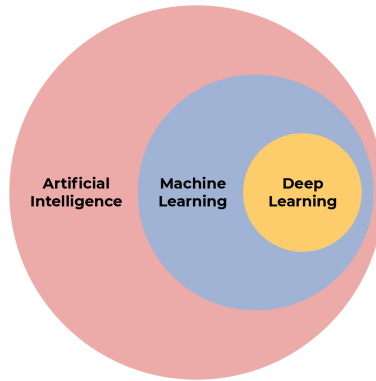


Figure 3: Diagramme de l'intelligence artificielle

Le Machine Learning peut être classé en trois grandes catégories :

- **Apprentissage supervisé**
- **Apprentissage non supervisé**
- **Apprentissage avec renforcement**

13.1 Apprentissage supervisé

Le machine learning supervisé est une technologie élémentaire mais stricte. Les opérateurs présentent à l'ordinateur des exemples d'entrées et les sorties souhaitées, et l'ordinateur recherche des solutions pour obtenir ces sorties en fonction de ces entrées. Le but est que l'ordinateur apprenne la règle générale qui mappe les entrées et les sorties. Le machine learning supervisé peut être utilisé pour faire des prédictions sur des données indisponibles ou futures (on parle alors de "modélisation prédictive"). L'algorithme essaie de développer une fonction qui prédit avec précision la sortie à partir des variables d'entrée. Le machine learning supervisé peut se subdiviser en deux types : Classification : la variable de sortie est une catégorie. Régression : la variable de sortie est une valeur spécifique. Les principaux algorithmes du machine learning supervisé sont les suivants : forêts aléatoires, arbres de décision, algorithme K-NN (k-Nearest Neighbors), régression linéaire, algorithme de Naïve Bayes, machine à vecteurs de support (SVM), régression logistique et boosting de gradient.

13.2 Apprentissage non supervisé

Dans le cadre du machine learning non supervisé, l'algorithme détermine lui-même la structure de l'entrée (aucune étiquette n'est appliquée à l'algorithme). Cette approche peut être un but en soi (qui permet de découvrir des structures enfouies dans les données) ou un moyen d'atteindre un certain but. Il existe deux types de machine learning non supervisé :

- **Clustering** : l'objectif consiste à trouver des regroupements dans les données.
- **Association** : l'objectif consiste à identifier les règles qui permettront de définir de grands groupes de données. Les principaux algorithmes du machine learning non supervisé sont les suivants : K-Means, clustering/regroupement hiérarchique et réduction de la dimensionnalité.

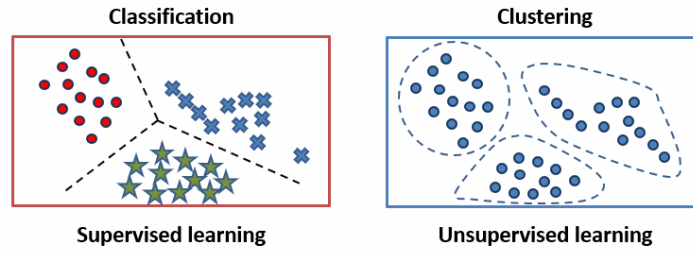


Figure 4: Comparaison apprentissage supervisé et non supervisé

13.3 Apprentissage avec renforcement

L'apprentissage par renforcement ou Reinforcement Learning est une méthode de Machine Learning. Elle consiste à entraîner des modèles d'intelligence artificielle d'une manière bien spécifique. L'agent IA doit apprendre à atteindre un objectif au sein d'un environnement incertain et potentiellement complexe. Pour y parvenir, l'ordinateur essaye toutes les façons possibles et apprend de ses erreurs. À chaque tentative, l'IA reçoit une récompense ou une punition en fonction des actions effectuées. Elle est programmée pour maximiser sa récompense, et tentera donc de trouver la méthode le lui permettant.

14 Méthodes de Prédiction du Diabète Utilisant le Machine Learning

- **Régression logistique (LR)** : Utilisée par Maniruzzaman , avec une précision de 90,62
- **Random Forest** : Adoptée par VijiyaKumar , avec une précision élevée.
- **Algorithme C4.5 DT** : Utilisé par Woldemichael , obtenant une précision élevée.
- **Réseaux de neurones artificiels (RNA)** : Utilisés par Benbelouaer , avec une précision de 83,33
- **K-Nearest Neighbours (KNN)** : Utilisé par Daanouni , avec une précision de 87,36
- **Réseaux de neurones convolutifs (D-CNN)** : Adoptés par Nareshkumar , avec une précision globale de 96,25
- **Algorithme de gradient boosting** : Performant selon Parisa , avec une précision de 95,50
- **Réseaux de neurones artificiels (ANN)** : Utilisés par Talha, avec une précision de 75,7
- **Random Forest** : Appliqué par Sidahmad , avec des taux de classification allant jusqu'à 91
- **Support Vector Machine (SVM) et K-Nearest Neighbours (KNN)** : Les plus appropriés selon Nasir , avec une précision de prédiction de 77

15 Méthode qui sera utilisée K-means

K-means est un algorithme de quantification vectorielle, également connu sous le nom de clustering. Son objectif est de séparer un ensemble de points en K clusters.

15.1 Notations, Mesure de distorsion

Les notations utilisées sont les suivantes :

- $\mathbf{x}_i \in \mathbb{R}^p, i \in \{1, \dots, n\}$ représentent les points à séparer.
- Les z_{ki} sont des variables indicatrices associées aux \mathbf{x}_i , telles que $z_{ki} = 1$ si \mathbf{x}_i appartient au cluster k , et $z_{ki} = 0$ sinon. \mathbf{z} est la matrice des z_{ki} .
- μ est le vecteur des $\mu_k \in \mathbb{R}^p$, où μ_k est le centre du cluster k .

La mesure de distorsion $J(\mu, \mathbf{z})$ est définie comme :

$$J(\mu, \mathbf{z}) = \sum_{i=1}^n \sum_{k=1}^n z_{ki} \|\mathbf{x}_i - \mu_k\|^2$$

15.2 Algorithme

L'algorithme de K-means vise à minimiser $J(\mu, \mathbf{z})$ et se présente sous la forme d'un algorithme de minimisation alternée :

1. Étape 0 : Choisir le vecteur μ .
2. Étape 1 : Minimiser J par rapport à \mathbf{z} : $z_{ki} = 1$ pour $k \in \operatorname{argmin} \|\mathbf{x}_i - \mu_k\|$.
3. Étape 2 : Minimiser J par rapport à μ : $\mu_k = \frac{\sum_i z_{ki} \mathbf{x}_i}{\sum_i z_{ki}}$.
4. Étape 3 : Retourner à l'étape 1 jusqu'à convergence.

15.3 Convergence et initialisation

Il est démontré que cet algorithme converge en un nombre fini d'opérations. Cependant, la convergence est locale, ce qui soulève le problème de l'initialisation. Une méthode classique consiste à lancer l'algorithme plusieurs fois avec des initialisations aléatoires des μ_k et à choisir la répartition avec la distorsion minimale.

15.4 Choix de K

Le choix de K n'est pas universel. En augmentant K , la distorsion diminue, mais il peut conduire à un sur-ajustement. Il est possible d'ajouter un terme dépendant de K dans J , mais ce choix reste arbitraire.

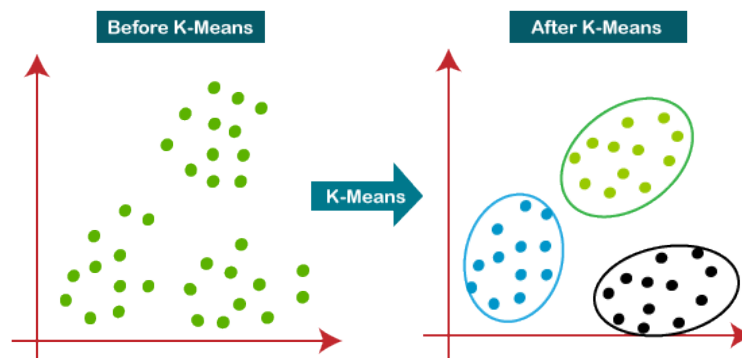


Figure 5: Schéma de représentation de k mean

16 Règles et contraintes mise en place

Pour garantir la précision et la fiabilité des données généalogiques, il est essentiel de définir des règles et des contraintes techniques spécifiques. Voici les principales règles à suivre pour assurer la cohérence de l'arbre généalogique :

16.1 Vérification des données

- Validation des Relations Familiales :
 - Assurer que les relations parent-enfant sont logiquement correctes, par exemple, un individu ne peut pas être son propre parent ou grand-parent.
 - Vérifier les dates de naissance et de décès pour garantir qu'un enfant ne peut pas naître avant ses parents ou après leur décès (sauf si le père décède, marge de 9 mois possible).
- Détection des Doublons :
 - Mettre en place des algorithmes pour détecter et signaler les doublons potentiels dans les individus et les relations, en utilisant des critères comme le nom, la date de naissance et les liens familiaux.

16.2 Intégrité des Données

- Contrôle de l'Unicité des Identifiants :
 - Assurer que chaque individu et chaque relation dans l'arbre possède un identifiant unique pour éviter les conflits et les ambiguïtés.
- Cohérence des Données Chronologiques :
 - Mettre en place des vérifications pour garantir que les événements chronologiques (naissance, mariage, décès) suivent une séquence logique et cohérente.

16.3 Gestion des Anomalies

- Gestion des Relations Inverses :
 - Lorsqu'une relation est ajoutée (par exemple, un parent à un enfant), s'assurer que la relation inverse est automatiquement créée et mise à jour (l'enfant vers le parent).
- Traitement des Cas Spéciaux :
 - Gérer les cas de familles recomposées, adoptions, et autres situations complexes en définissant clairement les types de relations et en les traitant de manière appropriée dans l'arbre généalogique.

16.4 Maintenance et Mise à Jour

- Historique des Modifications :
 - Implémenter un système de suivi des modifications pour enregistrer toutes les modifications apportées à l'arbre généalogique, permettant de revenir en arrière en cas d'erreur ou d'incohérence.
- Synchronisation des Données :
 - Assurer que les données importées à partir de fichiers CSV ou d'autres sources externes sont synchronisées correctement avec l'arbre existant, en respectant les règles de cohérence définies.

16.5 Contrainte de Performance

- Optimisation des Requêtes :
 - Utiliser des index et des requêtes optimisées pour gérer efficacement la recherche et la mise à jour des relations dans l'arbre généalogique, particulièrement pour les arbres de grande taille.
- Scalabilité :
 - Concevoir le système pour qu'il puisse évoluer afin de gérer un nombre croissant de relations et d'individus sans compromettre la performance ou l'intégrité des données.

En respectant ces règles et contraintes techniques, nous nous assurons que l'arbre généalogique reste cohérent, précis et fiable, permettant ainsi aux utilisateurs de bénéficier d'une expérience sans faille et de données généalogiques de haute qualité.

17 Conclusion

Nous avons bien défini tous les paramètres rentrant en jeu afin de créer un arbre généalogique quelconque cohérent; tout cela permis par un cahier des charges complet. Nous avons également évoqué les outils que nous allons utiliser dans la suite de notre projet à savoir un fichier CSV, un diagramme UML ou encore une implémentation Java.

D'autre part, nous nous sommes intéressés à une étude de cas bien précise: la prédiction du diabète. Nous avons donc introduit le diabète ainsi que son fonctionnement héréditaire dans un premier temps, puis nous avons défini le modèle mathématique et technologique à utiliser. Grâce à ce premier rendu, nous avons désormais une base solide et nous allons pouvoir, dans un second temps, réaliser un diagramme UML de notre problème afin de pouvoir effectuer une implémentation Java puis une interface graphique.

18 Source

- Jones, S. (2020). "Comparing the Top Genealogy Software Programs." Family History Daily.
- Smith, J. (2019). "10 Best Family Tree Builders of 2020." Top10.com
- Nabila YAYA Melissa LACHI.(2021-2022)."Prédiction du Diabète gestationnel en utilisant les techniques de l'intelligence artificielle"