

Assur Tech : Vers Un Système De Recommandation Pour Une Entité D'assurance Innovante

Laaroussi Ayoub , Ouchekkir Ali , Ouzian Rabab

¹faculty Of Legal, Economic And Social Sciences Agdal, Mohamed V University, Rabat, Morocco

²mohamed V University– Rabat, Morocco

³faculty Faculty Of Legal, Economic, And Social Sciences - Souissi, Mohammed V University, Rabat, Morocco

Résumé

Le marché des particuliers et des professionnels dénonce une incohérence perfectible entre les profils des clients et des usagers avec les produits souscrits. Cette incohérence se traduit par un manque à gagner pour l'assurance en termes de commissions et par un risque de perte du client en cas d'insatisfaction vis-à-vis du produit détenu. Dans cette optique, le Cross Selling constitue un véritable levier de performance à travers la recommandation au bon moment d'un produit et d'un service au bon client. Il s'agit alors de recommander au client les produits dont il pourrait avoir besoin pour augmenter sa rentabilité, en tenant compte de ses produits, son besoin, et son historique suivant une approche de similarité entre les clients ou une approche d'appétence à un produit.

Mots clés : AssurTech, technologie, Insurance solutions, Big Data Analytics, Intelligence Artificielle.

Date of Submission: 14-04-2024

Date of Acceptance: 24-04-2024

I. Introduction Générale

La croissance fulgurante de l'AssurTech a instauré une ère de transformation pour l'industrie de l'assurance, capitalisant sur les avancées technologiques et l'innovation pour accroître l'efficacité des services assurantiels. Selon la définition de l'Autorité de contrôle prudentiel et de résolution (Banque de France), l'AssurTech englobe les compagnies d'assurance, les intermédiaires ou les spécialistes de la chaîne de valeur de l'assurance qui recourent à la technologie pour rivaliser avec le secteur de l'assurance ou lui apporter une valeur ajoutée. Des entreprises novatrices telles que Lemonade, Trov et ZhongAn ont adopté des technologies comme l'intelligence artificielle et les chatbots pour rationaliser les processus de souscription et de règlement des sinistres.

Actuellement, l'AssurTech persévère dans sa trajectoire innovante en intégrant des systèmes de recommandation visant à renforcer les relations client et à soutenir le cross-selling, s'inscrivant ainsi dans la dynamique de la digitalisation croissante de l'économie. Cette dynamique expose un défi fondamental pour l'industrie : la personnalisation des offres d'assurance pour répondre de manière unique à chaque assuré.

Dans ce contexte évolutif, l'industrie de l'assurance est confrontée à une problématique fondamentale: la nécessité de personnaliser les offres pour répondre aux besoins spécifiques de chaque assuré. En réponse, les systèmes de recommandation se profilent comme des outils stratégiques, capables de cibler précisément les attentes individuelles et d'optimiser les relations client. Soulignant l'importance de la personnalisation, des études telles que celle de Deloitte (2024) mettent en avant que des offres d'assurance sur mesure améliorent non seulement la satisfaction des clients, mais contribuent également à accroître la fidélisation de la clientèle. Alors que les attentes des consommateurs évoluent rapidement, les compagnies d'assurance recherchent activement des moyens novateurs pour rester compétitives et pertinentes. Dans cette optique, l'étude vise à définir une méthodologie systématique pour le développement d'un système de recommandation adapté au secteur de l'assurance, mettant un accent particulier sur la personnalisation des produits et services. Le choix de la méthode **Cross Industry Standard Process for Data Mining** repose sur son succès avéré dans divers domaines, fournissant une structure méthodique pour la gestion du processus de découverte de connaissances à partir des données.

II. Revue De La Littérature

Systèmes de Recommandation dans le Contexte de l'Assurance

Le secteur de l'assurance est confronté à un paysage en constante évolution, marqué par des demandes croissantes de personnalisation et d'efficacité dans la prestation de services. Dans ce contexte dynamique, les systèmes de recommandation émergent comme des outils stratégiques permettant d'optimiser les interactions entre les assureurs et leurs clients. Ces systèmes, basés sur des techniques avancées d'exploration de données, offrent la possibilité de proposer des produits et des services adaptés aux besoins spécifiques de chaque assuré (Barbara et al., 2017).

Les systèmes de recommandation dans le domaine de l'assurance tirent parti des algorithmes sophistiqués

pour analyser les données des clients, tenant compte de leurs antécédents en matière de réclamations, de leurs préférences personnelles et de leurs comportements d'achat. Ces informations permettent aux assureurs de comprendre de manière approfondie les besoins individuels des clients et de recommander des polices d'assurance qui correspondent le mieux à leurs profils). L'efficacité des systèmes de recommandation dans le secteur de l'assurance repose sur leur capacité à évoluer en temps réel en fonction des changements dans la vie des assurés (Bittini JS, Rambaud SC, Pascual JL, Moro-Visconti R ,2022).

Intelligence Artificielle et Apprentissage Automatique

L'apprentissage automatique est un sous-domaine de l'intelligence artificielle (IA) qui se concentre sur la conception de systèmes capables d'apprendre ou d'améliorer leur performance en fonction des données qu'ils consomment. L'intelligence artificielle, en général, englobe des systèmes ou des machines qui imitent l'intelligence humaine.

Les termes "apprentissage automatique" et "intelligence artificielle" sont souvent associés et parfois confondus, mais ils ne sont pas identiques. L'apprentissage automatique est une composante de l'intelligence artificielle, qui couvre un champ beaucoup plus large. Aujourd'hui, l'apprentissage automatique est omniprésent dans notre quotidien. Lorsque nous interagissons avec des banques, des compagnies d'assurance, ou utilisons les réseaux sociaux, les algorithmes d'apprentissage automatique entrent en jeu pour rendre notre expérience plus fluide, efficace et sécuritaire. Les progrès dans l'apprentissage automatique et les technologies associées se développent rapidement, et nous commençons tout juste à entrevoir l'étendue de leurs capacités.

Système de recommandation : Définition et fonctionnement

Les systèmes de recommandation sont une forme spécifique de filtrage de l'information (SI) visant à présenter les éléments qui sont susceptibles d'intéresser l'utilisateur. Généralement, un système de recommandation permet de comparer le profil d'un utilisateur à certaines caractéristiques de référence, et cherche à prédire « l'avis » que donnerait un utilisateur. L'objectif principal des systèmes de recommandation est de fournir des suggestions pertinentes aux utilisateurs en ligne pour prendre de meilleures décisions à partir de nombreuses alternatives disponibles sur le Web. Un meilleur conseiller système est plus orienté vers des recommandations personnalisées en prenant en considération l'empreinte numérique disponible de l'utilisateur, telle que les informations démographiques détaillées, journaux d'interaction et informations sur un produit, tels que spécifications, commentaires des utilisateurs, comparaison avec d'autres produits, etc. avant de faire des recommandations (Charlin, L. (2017).

Construire un bon système de recommandation pose des défis à la fois aux acteurs du système, à savoir les consommateurs et les vendeurs. Du point de vue du consommateur, recevoir des suggestions pertinentes d'une source fiable est essentiel à la prise de décision, le moteur de recommandation doit donc être construit de manière à acquérir la confiance des consommateurs. Du point de vue du vendeur, il est plus important de générer des recommandations pertinentes pour les consommateurs à un niveau personnalisé. Avec la montée des ventes en ligne, les grands acteurs collectent maintenant de grands volumes de journaux d'interaction transactionnels des utilisateurs pour analyser les comportements des utilisateurs plus profondément que jamais. À l'égard des progrès de la technologie et de la recherche, les moteurs de recommandation évoluent pour surmonter ces défis basés sur l'analyse de données volumineuses et l'intelligence artificielle (Su et Khoshgoftaar, 2009).

Le Big Data et les Systèmes de Recommandation

Récemment, les systèmes de recommandation ont réussi à avoir un impact sur nos vies de plusieurs façons. Un exemple évident de cet impact est la manière dont notre expérience d'achat en ligne a été redéfinie. Lorsque nous naviguons sur des sites de commerce électronique et que nous achetons des produits, les moteurs de recommandation répondent immédiatement, en temps réel, aux diverses suggestions pertinentes aux consommateurs. Indépendamment de la perspective, d'un consommateur, les systèmes de recommandation ont été extrêmement bénéfiques. Sans aucun doute, le big data est la force motrice des systèmes de recommandation (Gandomi, A., & Haider, M. (2015).

Un système de recommandation typique ne peut pas faire son travail efficacement sans données suffisantes. L'introduction de la technologie Big Data a permis aux entreprises de capturer de nombreuses données utilisateur, telles que les achats antérieurs, l'historique de navigation et les Informations de retour, et de les transmettre aux systèmes de recommandations pour générer des recommandations pertinentes et efficaces en temps réel. En bref, même le système de recommandation le plus avancé ne peut pas être efficace sans la fourniture de gros volumes de données. Le rôle du Big Data et des améliorations technologiques, tant sur le plan logiciel que matériel, va au-delà de la simple fourniture de données massives. Il fournit également des données significatives et exploitables rapidement, et fournit la configuration nécessaire pour traiter rapidement les données en temps réel.

Les approches des systèmes de recommandation

Maintenant que nous avons défini les systèmes de recommandation, leur objectif, leur utilité et la force motrice derrière les systèmes de recommandation, nous présentons dans cette section les différents types de systèmes de recommandation les plus populaires.

Filtrage collaboratif

Les systèmes de recommandation du filtrage collaboratif sont des formes de base de moteurs de recommandation. Dans ce type de moteur, le filtrage des éléments d'un grand nombre d'alternatives est effectué en collaboration par les préférences des utilisateurs. L'hypothèse de base dans un système de recommandation de filtrage collaboratif est que si deux utilisateurs partageaient les mêmes intérêts les uns par rapport aux autres dans le passé, ils auront également des goûts similaires dans le futur. Si, par exemple, l'utilisateur A et l'utilisateur B ont des préférences de film similaires et l'utilisateur A avait récemment regardé Titanic, que l'utilisateur B n'a pas encore vu, alors l'idée est de recommander ce nouveau film à l'utilisateur B. Les recommandations de films sur Netflix sont un bon exemple de ce type de système de recommandation.

→ Il existe deux types de systèmes de recommandation de filtrage collaboratif :

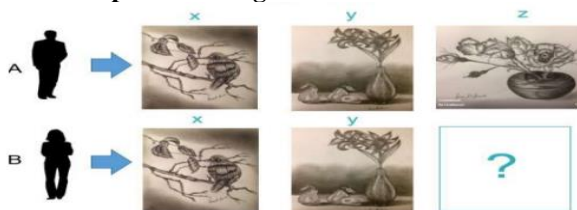
❖ Filtrage collaboratif basé sur les utilisateurs (U2U) :

Les recommandations sont générées en considérant les préférences des utilisateurs voisins. Le filtrage collaboratif basé sur l'utilisateur se fait en deux étapes :

- Identifier des utilisateurs similaires en fonction des préférences des utilisateurs
- Recommander de nouveaux éléments à un utilisateur actif en fonction de la note donnée par des utilisateurs similaires sur les articles non notés par l'utilisateur actif.

Exemple : Si un client A clique sur la photo x, y et z, et la personne B clique sur la photo x et y, le système de recommandation va recommander la photo z à la personne B.

Figure 1 : Exemple du Filtrage collaboratif basé sur les utilisateurs



Source : élaborer par nous-même

❖ Filtrage collaborative basé sur les items (I2I) :

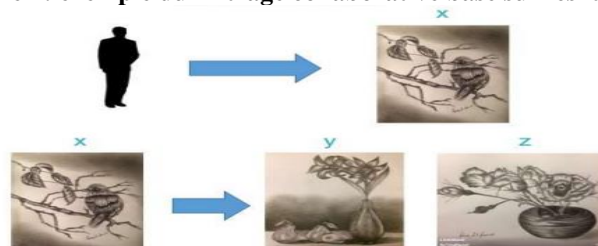
Les recommandations sont générées en utilisant le voisinage des produits. Contrairement au filtrage collaboratif basé sur l'utilisateur, nous trouvons d'abord des similitudes entre les items, puis nous recommandons des items non notés qui sont similaires aux éléments que l'utilisateur actif a notés dans le passé. Les systèmes de recommandation basés sur les items sont construits en deux étapes :

- Calculer la similarité des items en fonction des préférences de l'item.
- Trouver les meilleurs items similaires aux items non évalués par l'utilisateur actif et les recommander.

Lorsque nous appliquons une similarité de cosinus aux systèmes de recommandation, nous considérons la colonne d'élément comme le vecteur à n dimensions et la similarité entre deux éléments comme l'angle entre eux. Plus l'angle est petit, plus les éléments sont similaires.

Exemple : Lorsque la personne clique sur tableau x, le système de recommandation recommande les tableaux similaires de x qui sont dans cette exemple est y et z.

Figure 2: exemple du Filtrage collaborative basé sur les items



Source : élaborer par nous-même

Filtrage basé sur le contenu

Les approches basées sur le contenu, appelées également filtrage par le contenu, sont les premières approches de recommandation à avoir vu le jour. Elles consistent à recommander à l'utilisateur des items similaires, du point de vue de leur contenu, aux items qu'il aura appréciés auparavant. Le filtrage basé sur le contenu nécessite de disposer d'une description des items à recommander et d'une connaissance des préférences de l'utilisateur.

Les approches basées sur le contenu dérivent du monde de la recherche d'informations. C'est pourquoi les premiers systèmes ayant été mis en place étaient le plus souvent appliqués dans des systèmes d'information à contenu textuel. Ceci afin de pouvoir utiliser les techniques existantes dans le domaine de la recherche d'information. Ces recommandations se basent donc sur des mots-clés décrivant les items. Ces mots-clés peuvent être pondérés afin d'avoir une représentation plus précise de chaque item. L'une des mesures les plus utilisées pour exécuter cette pondération est le TF-ID. Ce dernier peut être utilisé pour améliorer la pertinence des recommandations de produits d'assurance en fonction des besoins et des caractéristiques spécifiques des clients.

Les mots-clés et leurs pondérations ainsi récupérées permettent de d'écrire les items sous forme de vecteurs de poids où chaque composante représente le poids des mots-clés, calculé grâce à TF-IDF par exemple. La mesure la plus utilisée pour le calcul de similarités dans cette configuration est la similarité cosinus.

$$sim = (d_j, d_k) = \frac{\vec{d}_j \cdot \vec{d}_k}{|\vec{d}_j| |\vec{d}_k|} = \frac{\sum_{i=1}^n W_{i,j} W_{i,k}}{\sqrt{\sum_{i=1}^n W_{i,j}^2} \sqrt{\sum_{i=1}^n W_{i,k}^2}}$$

La construction de systèmes de recommandation basés sur le contenu implique trois étapes principales, comme suit :

- Générer des informations de contenu pour les produits.
- Générer un profil d'utilisateur et des préférences en ce qui concerne les caractéristiques des produits.
- Générer des recommandations et prédire une liste d'éléments dont l'utilisateur pourrait avoir besoin.

Ces deux approches (approche basée sur le contenu et approche basée sur le filtrage collaboratif) ne sont pas disjointes et sont la plupart du temps combinées et construisent une autre approche qui s'appelle « approche hybride ».

Les approches hybrides

Les approches hybrides sont des approches qui combinent deux ou plusieurs approches de recommandation. Ces combinaisons permettent de bénéficier des avantages des approches utilisées, de pallier leurs inconvénients et de proposer des recommandations plus pertinentes. Par exemple, le démarrage à froid au niveau des items des approches collaboratives, peut bénéficier des avantages des approches basées sur le contenu. En effet, dans une approche basée sur le contenu, seule la description de l'item est utilisée dans la recommandation, il n'y a donc pas besoin que l'item soit noté par un certain nombre d'utilisateurs avant de pouvoir être recommandé, comme c'est le cas dans les approches collaboratives.

Autres types de systèmes de recommandation

Type	Méthode utilisée
Les approches démographiques	Dans une approche démographique, le système de recommandation demande à l'utilisateur de renseigner un certain nombre de paramètres (âge, sexe, ville, intérêts, etc.). Ces paramètres permettent de définir des classes démographiques dans lesquelles les utilisateurs sont catégorisés. Par la suite, les recommandations de chaque utilisateur proviennent de la classe qui lui a été attribuée.
Les approches sociales	Les réseaux sociaux introduisent un nouveau paramètre dans la recommandation en exploitant le cercle social de l'utilisateur. Ainsi, un système de recommandation social peut suggérer des éléments populaires au sein du cercle social de l'utilisateur, adoptant une approche de filtrage collaboratif U2U. Cette méthode remplace le calcul traditionnel du voisinage de l'utilisateur par l'utilisation de son cercle social, résolvant ainsi le défi d'évolutivité. L'objectif est d'évaluer la confiance de l'utilisateur envers ses pairs pour identifier ceux qui sont plus similaires à son profil. Enfin, les items les plus populaires au sein du groupe identifié sont recommandés à l'utilisateur cible.
Les approches basées sur la connaissance	Le mode opératoire dans ce genre de système est conversationnel et l'utilisateur est amené à définir à chaque fois un certain nombre de critères pouvant mieux répondre à ses attentes. Il ne s'agit donc pas de construire un profil utilisateur à long terme mais de demander systématiquement à l'utilisateur de déterminer ses besoins explicitement. Le système cherche donc les items remplissant au mieux les conditions définies par l'utilisateur pour les lui présenter. Ceci peut être effectué grâce à un ensemble de règles définies par l'expert qui filtrent les items selon les critères définis par l'utilisateur.

Source : Concevoir de nos propres mains

Data Mining pour la prédiction

Définition et objectifs du data mining

Le data mining, également connu sous le nom d'exploration de données ou de fouille de données, est le processus d'extraction de connaissances ou de modèles précieux à partir de grandes quantités de données. Il fait partie du domaine plus large de l'analyse de données.

L'objectif principal du data mining est de découvrir des modèles, des relations ou des informations cachées à partir des données, qui peuvent être utilisés pour prendre des décisions éclairées, prédire des résultats futurs, optimiser des processus ou identifier des tendances. Le data mining utilise des techniques et des algorithmes statistiques, mathématiques et d'apprentissage automatique pour extraire ces connaissances.

Les objectifs spécifiques du data mining peuvent varier en fonction des besoins et des domaines d'application. Voici quelques objectifs courants du data mining:

- **Prédiction** : Utiliser des modèles et des algorithmes pour prédire des valeurs ou des événements futurs. Par exemple, prédire les ventes d'un produit en fonction des données historiques.
- **Classification** : Regrouper les données dans des catégories ou des classes spécifiques en fonction de leurs caractéristiques communes. Par exemple, classer les clients en segments pour une campagne marketing ciblée.
- **Regroupement (clustering)** : Détecter des similarités et des motifs intrinsèques dans les données pour former des groupes homogènes. Par exemple, regrouper les clients en segments basés sur leurs comportements d'achat.
- **Détection d'anomalies** : Identifier des schémas ou des observations inhabituelles dans les données qui peuvent être des anomalies ou des valeurs aberrantes. Par exemple, détecter des transactions frauduleuses dans les données financières.

En résumé, le data mining est un processus d'extraction de connaissances utiles à partir de grandes quantités de données, avec pour objectif de prendre des décisions éclairées, de prédire des résultats futurs, d'optimiser des processus et d'identifier des tendances

Rôle du data mining dans les prédictions temporelles : Le cas des assurances

Le data mining joue un rôle important dans les prédictions temporelles en exploitant les modèles et les relations cachés dans les données temporelles (Park BK, Yeo SH, Lee S, Yang HD,2020).

Voici quelques façons dont le data mining est utilisé pour les prédictions temporelles dans le secteur de l'assurance :

- **Analyse des tendances** : Les modèles de séries chronologiques appliqués au data mining se révèlent pertinents pour examiner les données temporelles spécifiques à l'industrie de l'assurance. En utilisant des algorithmes tels que la régression linéaire, les méthodes de lissage exponentiel ou les modèles ARIMA, le data mining peut élaborer des modèles prédictifs permettant d'anticiper les évolutions futures des sinistres, des primes et d'autres variables majeures.
- **Prédiction des tendances** : En analysant les données historiques propres à l'assurance, le data mining peut détecter les tendances émergentes et anticiper leur développement futur. Par exemple, dans le domaine de l'assurance automobile, le data mining pourrait être employé pour prévoir les tendances liées aux sinistres et aux réclamations, permettant ainsi aux assureurs d'ajuster leurs stratégies et leurs tarifs.
- **Prédiction des valeurs futures** : Le data mining offre la possibilité de prédire les valeurs futures de variables spécifiques, comme les indemnités de sinistre, en utilisant des modèles prédictifs tels que la régression, les réseaux neuronaux ou les machines à vecteurs de support (SVM). Ces modèles, en tenant compte des données temporelles antérieures et des caractéristiques associées, peuvent jouer un rôle essentiel dans l'estimation des coûts futurs pour les compagnies d'assurance.
- **Prévision de la demande** : Dans le domaine de l'assurance, le data mining peut être déployé pour anticiper la demande future de certaines couvertures en se basant sur des données historiques, les saisons, les événements particuliers et d'autres variables pertinentes. Cela permet aux assureurs d'ajuster leurs offres et leurs capacités opérationnelles pour répondre de manière proactive à la demande anticipée, améliorant ainsi leur efficacité globale.

III. Méthodologie

Problématique

La problématique soulevée constitue un enjeu majeur au sein de l'industrie de l'assurance, mettant en lumière la nécessité d'explorer les possibilités offertes par les systèmes de recommandation, en particulier lorsqu'ils sont guidés par la méthodologie CRISP (Cross Industry Standard Process for Data Mining) (Halima EH, Yassine T,2022). Cette méthodologie, reconnue pour son approche systématique en matière de gestion de projets d'exploration de données, offre un cadre structuré pour aborder le défi majeur de la personnalisation des offres au sein du secteur de l'assurance.

Dans un environnement où les attentes des clients évoluent rapidement, la personnalisation est devenue

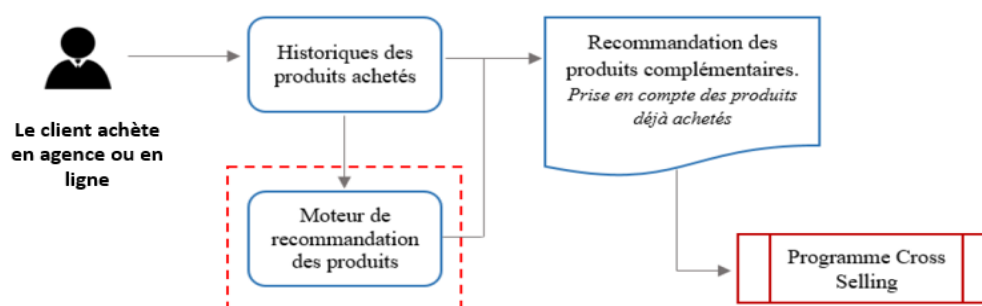
essentiel pour les compagnies d'assurance, jouant un rôle clé dans leur adaptabilité et leur succès. Les systèmes de recommandation sont essentiels pour répondre à ce défi en analysant efficacement les données complexes, en identifiant les tendances spécifiques et en fournissant des recommandations individualisées. La personnalisation des offres dépasse la simple adaptation des produits d'assurance ; elle optimise également les relations client et améliore la performance globale des compagnies d'assurance (Cappiello A,2018). En renforçant la fidélisation, en accroissant la satisfaction client et en favorisant la rétention, elle maintient les compagnies compétitives dans un environnement en constante évolution où l'innovation et la différenciation sont cruciales(Cappiello A,2020).

Ainsi, la recherche sur la manière dont les systèmes de recommandation, suivant la méthodologie CRISP, peuvent répondre au défi de la personnalisation des offres dans l'industrie de l'assurance est essentielle pour anticiper les tendances du marché, répondre aux attentes des clients et positionner les compagnies d'assurance comme des acteurs agiles et visionnaires. Cette problématique offre une opportunité unique d'explorer les synergies entre la technologie des recommandations, la méthodologie CRISP et les impératifs stratégiques des compagnies d'assurance dans un contexte de mutation rapide (Chang VY, 2023). Dans ce contexte, nous formulons les hypothèses suivantes :

Hypothèse	Formulation	Justification
H1	Similarité entre les profils clients et succès des recommandations	Cette hypothèse suggère qu'une similarité entre les profils clients contribue positivement au succès des recommandations. La similarité entre les profils pourrait également influencer l'appétence à un produit, créant ainsi un lien entre la similarité et les choix de recommandations. Une corrélation potentielle existe entre la similarité des profils et la perception positive des clients vis-à-vis des produits recommandés.
H2	Personnalisation des recommandations et adéquation aux besoins clients	L'hypothèse sur la personnalisation suggère que des recommandations personnalisées influencent positivement le succès global des recommandations. La personnalisation des recommandations est également liée à une augmentation de la satisfaction client.
H3	Appétence à un produit et succès du Cross Selling	L'appétence à un produit peut être liée à la similarité entre les profils clients, créant ainsi une connexion entre les préférences individuelles et la similarité des profils. L'influence de l'appétence à un produit sur le choix des recommandations est mise en évidence.
H4	Moment opportun et taux de succès des recommandations	Le choix du moment opportun pour les recommandations a un impact positif sur la personnalisation des recommandations. Il peut également être lié à la similarité entre les profils clients, soulignant l'importance du timing en relation avec les caractéristiques communes des clients.

Source : Concevoir de nos propres mains

Figure 3 : Architecture globale du modèle de prédiction

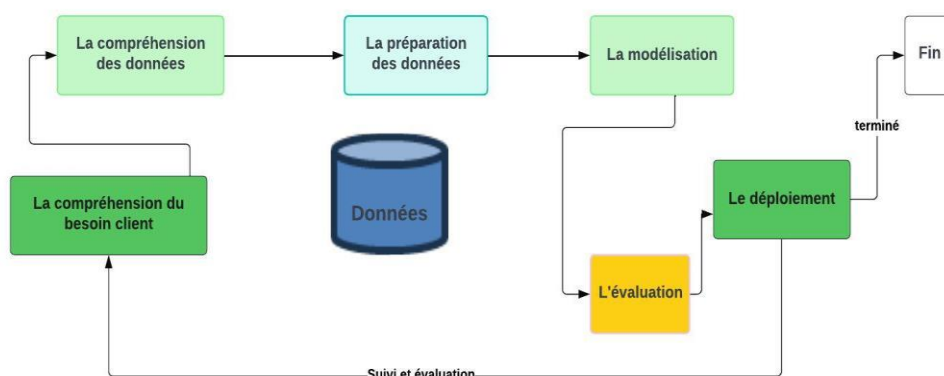


Source : élaborer par nous-même

Méthode de gestion de projet

La méthodologie suivie est la méthode CRISP (Cross Industry Standard Process for Data Mining), cette méthode est utilisée pour mener des projets de data mining. Elle fournit une structure et des étapes clés pour guider le processus de data mining, de la compréhension des besoins commerciaux à la mise en œuvre des modèles prédictifs.

Figure 4 : La méthodologie CRISP



Source : élaborer par nous-même

❖ **Étude & analyse :**

- Comprendre le processus actuel de souscription d'assurance de A à Z.
- Identifier les points d'interaction avec les clients et les opportunités d'amélioration.
- Proposer un nouveau processus intégrant un système de recommandation au sein d'une compagnie d'assurance

❖ **Collecte et préparation des données :**

- Identifier les sources de données pertinentes, telles que les historiques de souscription, les caractéristiques des assurés, les réclamations passées, etc.
- Veiller à ce que les données soient complètes, fiables et adaptées à l'analyse.

❖ **Analyse des données :**

- Effectuer une analyse approfondie des données pour comprendre les tendances et les schémas.
- Utiliser des techniques d'exploration de données pour identifier les facteurs influençant les choix d'assurance des clients.

❖ **Modèle de recommandation :**

- Développer un modèle de recommandation basé sur les données analysées.
- Utiliser des techniques de modélisation prédictive adaptées au domaine de l'assurance.
- Intégrer des variables telles que le profil de l'assuré, l'historique des réclamations, les préférences, etc.

❖ **Conception de la plateforme :**

- Concevoir l'architecture de la plateforme de recommandation d'assurance.
- Identifier les fonctionnalités nécessaires, telles que la personnalisation des offres, la gestion des polices, etc.
- Sélectionner les aspects technologiques, tels que les frameworks de développement et les langages de programmation adaptés.

❖ **Réalisation :**

- Mettre en œuvre la plateforme en utilisant un langage de programmation adapté.
- Intégrer le modèle de recommandation dans la plateforme pour fournir des recommandations en temps réel lors de la souscription d'une assurance.
- Assurer que la plateforme prend en charge efficacement la gestion des polices, la personnalisation des offres et l'affichage des prévisions.

En combinant l'analyse des données et un modèle de recommandation personnalisé, cette approche vise à optimiser le processus de souscription d'assurance en fournissant des recommandations pertinentes et adaptées aux besoins spécifiques de chaque assuré.

Benchmarking

Applications	Méthodes utilisées
Personnalisation des Polices d'Assurance	<ul style="list-style-type: none"> ▪ Filtrage Collaboratif basé sur la Mémoire ▪ Données : Historique des polices souscrites, préférences des assurés, caractéristiques des couvertures

Recommandations de Produits d'Assurance	<ul style="list-style-type: none"> ▪ Filtrage Collaboratif (Produits + Utilisateurs) ▪ Données : Profil démographique (identifiant client, âge, sexe, profession...), historique des transactions
Système de Gestion de Risques	<ul style="list-style-type: none"> ▪ Méthode Hybride (Filtrage Collaboratif + Filtrage basé sur le Contenu) ▪ Données : Historique des sinistres, données sur les risques couverts, préférences des assurés
Recommandations de Services Complémentaires	<ul style="list-style-type: none"> ▪ Filtrage Collaboratif basé sur la Mémoire ▪ Données : Historique des services souscrits, préférences des assurés, comportements d'utilisation

IV. Analyse Du Problème Et Solution

Mesures de similarité

Les techniques d'analyse de données sont au cœur des systèmes de recommandation. Ces techniques d'exploration de données nous aident à extraire des modèles, à regrouper des utilisateurs, à calculer des similarités, à prédire des préférences, à gérer des données d'entrée, à évaluer des modèles de recommandation, etc.

Cosinus similarité

La similarité cosinus (ou mesure cosinus) permet de calculer la similarité entre deux vecteurs à n dimensions en déterminant le cosinus de l'angle entre eux.

Il est défini par la formule suivante :

$$sim = (d_j, d_k) = \frac{\vec{d}_j \cdot \vec{d}_k}{|\vec{d}_j| |\vec{d}_k|} = \frac{\sum_{i=1}^n W_{ij} W_{ik}}{\sqrt{\sum_{i=1}^n W_{ij}^2} \sqrt{\sum_{i=1}^n W_{ik}^2}}$$

Puisque la valeur $\cos \theta$ se situe dans la plage de $[-1, 1]$, un score de -1 signifie que les vecteurs sont totalement opposés, 0 indique qu'ils sont indépendants (orthogonaux), et 1 suggère qu'ils sont similaires (colinéaires avec un coefficient positif). Les valeurs intermédiaires facilitent l'évaluation du degré de similitude entre les vecteurs.

Coefficient de corrélation de Pearson

Ce coefficient a été utilisé pour calculer la similarité entre deux utilisateurs u et v. Le coefficient de corrélation de Pearson mesure le rapport entre la covariance et le produit de l'écart-type des notes données par les deux utilisateurs. Il permet ainsi de mesurer la similarité en utilisant les items notés à la fois par u et v. Plus les deux utilisateurs auront tendance à noter les mêmes items de façon équivalente, plus ils seront similaires. Le coefficient de corrélation de Pearson peut également être utilisé pour mesurer la corrélation entre deux items i et j.

$$sim(i, j) = Pearson(i, j) = \frac{\sum_u (r_{u,i} - \bar{r}_i) \cdot (r_{u,j} - \bar{r}_j)}{\sqrt{\sum_{u \in u_{ij}} (r_{u,i} - \bar{r}_i)^2} \sqrt{\sum_{u \in u_{ij}} (r_{u,j} - \bar{r}_j)^2}}$$

Maintenant, comment le coefficient de corrélation de Pearson est-il appliqué dans les moteurs de recommandation ?

Considérons une matrice d'évaluation avec les IDs des utilisateurs en tant que lignes et les IDs des articles en tant que colonnes. Chaque ligne peut être considérée comme un vecteur utilisateur, tandis que chaque colonne représente un vecteur d'article. Le coefficient de corrélation entre les vecteurs de ligne offre une mesure de similarité entre les utilisateurs, tandis que le coefficient de corrélation entre les vecteurs de colonne représente la similarité entre les articles. Les valeurs obtenues se situent dans une plage de -1 à 1, où -1 indique une différence parfaite et 1 indique une similitude parfaite.

Distance euclidienne

En mathématiques, la distance euclidienne ou métrique euclidienne est la distance "ordinaire" en ligne droite entre deux points de l'espace euclidien. Avec cette distance, l'espace euclidien devient un espace métrique. La norme associée s'appelle la norme euclidienne. La littérature plus ancienne se réfère à la métrique en tant que métrique de Pythagore. Un terme généralisé pour la norme euclidienne est la norme L2 ou la distance L2.

La distance euclidienne entre les points p et q est la longueur du segment de ligne qui les relie en coordonnées cartésiennes, si $p = (p_1, p_2, \dots, p_n)$ et $q = (q_1, q_2, \dots, q_n)$ sont deux points de l'espace n euclidien, la distance (d) de p à q, ou de q à p est donné par la formule de Pythagore :

$$Distance\ euclidienne(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2}$$

L'utilisation de la distance euclidienne comme mesure de similarité dans les systèmes de

recommandation en assurance vise à évaluer la proximité entre les profils d'assurés ou les caractéristiques des contrats d'assurance. Cette approche peut être appliquée à diverses fins dans le domaine de l'assurance.

Évaluation des systèmes de recommandation

L'utilisation des systèmes de recommandation dans divers domaines a donné lieu à de nombreux algorithmes de recommandation. Comme le choix d'une méthode de recommandation appropriée pour un système donné nécessite la comparaison de plusieurs algorithmes, la communauté de recherche de la RS s'est attachée à l'étude des techniques d'évaluation efficace. De nombreux chercheurs évaluent les systèmes de recommandation à l'aide d'une analyse hors ligne, de méthodes expérimentales en direct (évaluation en ligne) ou d'une combinaison de ces deux méthodes.

Dans les méthodes d'évaluation en ligne, les utilisateurs réels interagissent avec un système de recommandation et reçoivent des recommandations qu'ils peuvent ou non accepter. La qualité de la recommandation peut ensuite être mesurée en demandant aux utilisateurs leur avis ou en surveillant leur comportement au cours de leur interaction avec le système (par exemple, en observant s'ils cliquent ou non sur les éléments recommandés). Les évaluations en ligne sont généralement coûteuses et prennent du temps (Kalika, M., & Kefi, H. 2004).

L'évaluation hors ligne

L'évaluation hors ligne, une technique répandue dans l'évaluation des systèmes de recommandation, repose sur la mesure de l'exactitude des recommandations sans la participation des utilisateurs réels. Cette approche utilise un ensemble de données existant comme vérité de terrain, divisant le jeu de données en un ensemble d'apprentissage et un ensemble de test. Le système de recommandation tente ensuite de prédire les évaluations dans l'ensemble de test en utilisant l'ensemble d'apprentissage, et les résultats sont comparés aux évaluations réelles dans l'ensemble de test. Parmi les métriques d'évaluation largement adoptées, les mesures de précision, telles que la précision et le rappel, évaluent à quel point les prédictions du système sont proches de la vérité au sol. L'erreur quadratique moyenne (RMSE) est également utilisée pour mesurer la concordance entre les cotes réelles des utilisateurs et les prédictions du système.

Ces mesures traditionnelles sont encore couramment utilisées, mais la communauté de la recherche en recommandation explore des paramètres d'évaluation plus holistiques pour couvrir divers aspects des recommandations. Des aspects tels que la couverture, la diversité et la nouveauté deviennent des critères d'évaluation pertinents. Ces paramètres offrent une vision plus complète de la performance des systèmes de recommandation, prenant en compte non seulement la précision, mais également des aspects cruciaux tels que la portée des recommandations, leur variété et leur capacité à introduire de nouvelles suggestions dans l'expérience utilisateur.

Les mesures de la qualité de l'évaluation hors ligne :

Précision	La précision est le nombre de documents pertinents retrouvés rapporté au nombre de documents total proposé par le moteur de recherche pour une requête donnée.
Rappel	Le rappel est défini par le nombre de documents pertinents retrouvés au regard du nombre de documents pertinents que possède la base de données.
RMSE	La déviation racine carrée ou erreur racine est une différence entre les valeurs prédites et les valeurs observées.
MAE	Erreur absolue moyenne (MAE) : MAE mesure la magnitude moyenne des erreurs dans un ensemble de prévisions, sans tenir compte de leur direction. C'est la moyenne sur l'échantillon test des différences absolues entre la prévision et l'observation réelle où toutes les différences individuelles ont un poids égal.
La diversité	On peut considérer la diversité comme une mesure de la diversité des éléments de système de recommandation (1-S) $S = (\text{similarité moyenne entre les paires de recommandations})$ Des scores de diversité inhabituellement élevés signifient qu'on a une mauvaise recommandation le plus souvent. Vous devez toujours rechercher la diversité aux côtés d'indicateurs qui mesurent également la qualité des recommandations.
Nouveauté	La nouveauté est une mesure de la popularité des objets que vous recommandez et encore une fois, le simple fait de recommander des objets aléatoires produirait des scores de nouveauté très élevés, car la grande majorité des objets ne sont pas très vendus.

Source : élaborer par nous-même.

Évaluation en ligne

Dans le domaine de l'évaluation des systèmes de recommandation, deux approches distinctes sont couramment utilisées : l'évaluation hors ligne, souvent réalisée à l'aide de la split-validation, et l'évaluation en ligne, qui privilégie le test A/B (ou test multivarié) comme méthode prédominante de nos jours. Alors que la split-validation est une méthode classique basée sur la division du jeu de données en ensembles d'apprentissage et de test, le test A/B en évaluation en ligne se démarque par son approche plus dynamique et orientée utilisateur.

L'évaluation en ligne implique l'intégration de plusieurs systèmes de recommandation (RS) différents, la division des utilisateurs en groupes distincts, et la confrontation des différentes solutions. Bien que cette approche puisse s'avérer coûteuse en termes de ressources de développement, elle offre une évaluation plus directe et en temps réel de la performance des recommandations. Cette méthode se révèle plus efficace par rapport à l'évaluation hors ligne, car elle prend en compte des indicateurs concrets tels que l'augmentation du nombre de clics, du bénéfice généré, du temps passé par les utilisateurs sur les pages, la satisfaction des utilisateurs, entre autres. Des études telles que celle menée par (Kalika, M., & Kefi, H. 2004) ont mis en évidence l'importance de l'évaluation en ligne pour mesurer l'impact réel des systèmes de recommandation sur les comportements des utilisateurs.

De plus, des recherches récentes ont mis en avant l'importance du test A/B dans l'évaluation en ligne en raison de sa capacité à produire des résultats robustes et instantanés. Ainsi, la validation en ligne, en mettant l'accent sur des indicateurs objectifs et concrets, offre une approche plus dynamique pour évaluer l'efficacité des systèmes de recommandation dans des conditions réelles d'utilisation.

Le Clustering

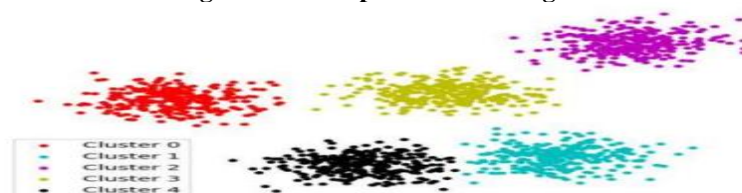
Le Clustering (partitionnement de données)

Le Clustering est une méthode en analyse des données. Elle vise à diviser un ensemble de données en différents « paquets » homogènes, en ce sens que les données de chaque sous-ensemble partagent des caractéristiques communes, qui correspondent le plus souvent à des critères de proximité (similarité informatique) que l'on définit en introduisant des mesures et classes de distance entre objets. Pour obtenir un bon partitionnement, il faut à la fois :

- Minimiser l'inertie intra-classe pour obtenir des grappes (cluster en anglais) les plus homogènes possibles ;
- Maximiser l'inertie interclasse afin d'obtenir des sous-ensembles bien différenciés.

Le clustering est une méthode d'apprentissage automatique qui regroupe des points de données en fonction de leur similarité ou de leur distance. Contrairement à la classification supervisée où les données sont déjà étiquetées, le clustering est non supervisé. Il s'agit d'une technique populaire en analyse statistique des données. Les algorithmes de classification peuvent être utilisés pour regrouper les points de données en groupes spécifiques, où les points d'un même groupe ont des propriétés similaires, tandis que ceux de groupes différents ont des caractéristiques distinctes (Bouveyron C, Celeux G, Murphy TB, & Raftery AE 2019).

Figure 5 : Exemple du Clustering



Source: Brightcape

Le clustering est utilisé notamment lorsqu'il est coûteux d'étiqueter les données. C'est néanmoins un problème mal défini mathématiquement : différentes métriques et/ou différentes représentations des données aboutiront à différents regroupements sans qu'aucun ne soit nécessairement meilleur qu'un autre. Ainsi la méthode de clustering doit être choisie avec soin en fonction du résultat attendu et de l'utilisation prévue des données.

Clustering par partitionnement

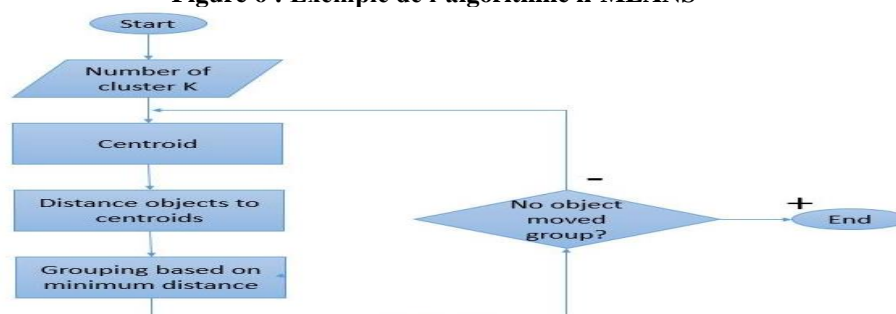
Les techniques par partitionnement créent un partitionnement des points de données, d'un seul niveau. Si k est le nombre désiré de clusters, alors les approches par partitionnement trouvent typiquement tous les k clusters immédiatement. Les techniques par partitionnement sont divisées en deux sous-catégories principales les algorithmes basés sur les centroïdes et les algorithmes basés sur les médoïdes.

- Pour Kmeans : on emploie la notion du centroïde qui est le point de la moyenne ou la médiane d'un groupe de points.
- Pour Kmedoid : on utilise la notion d'un médoïde qui est le point le plus représentatif (central) d'un groupe de points

Algorithme k-MEANS

La technique de clustering de Kmeans est très simple, son algorithme de base est décrit comme suit :

Figure 6 : Exemple de l'algorithme k-MEANS



Source : <https://researchhubs.com/post/ai/fundamentals/k-means-algorithm.html>

1. Choisir k comme nombre de cluster
2. Spécifier les centroides de chaque cluster.
3. Calcule de distance par rapport aux centroides.
4. Grouper les objets en se basant sur la distance minimale.
5. Si les centroides ont changé de position suite à la mise à jour, nous répétons les étapes 2, 3 et 4. Sinon, l'algorithme se termine et nous obtenons la partition finale des données en clusters

Algorithme k-Medoid

K-Medoid : Au lieu de prendre la valeur moyenne de l'objet dans un cluster comme point de référence, les medoids peuvent être utilisés, qui est l'objet le plus central dans un cluster. Dans l'approche K-medoid, un cluster est représenté par un de ses points. Ce point représentatif est appelé médoïde, c'est un point qui est le plus placé au centre en tenant en compte quelques mesures, comme par exemple, la distance. Chaque classe est représentée par un objet ou représentant de cette classe (medoid). Il s'agit aussi d'une méthode d'optimisation itérative combinant la réaffectation des points aux classes. Avec une intervention des représentants des points et des autres points d'où une stratégie couteuse et donc valable pour les données de petite taille.

Plusieurs algorithmes basés sur la notion de médoïde existent comme PAM (Partitioning Around Medoids) qui est un algorithme K-medoid qui essaye de grouper un ensemble de m points en k clusters en exécutant les étapes suivantes. Supposons qu'une entité d'assurance souhaite recommander des produits d'assurance à sa clientèle, en fonction de leurs besoins. L'entité dispose de données clients comprenant des informations telles que : l'âge, le sexe, les ayants droit, les bénéficiaires, etc.

V. Défis Actuels Dans La Recommandation De Produits D'assurance

Le filtrage collaboratif, technique prédominante en recommandation, repose sur l'utilisation des avis d'autres utilisateurs pour orienter les utilisateurs vers des produits et services. Cette approche, fondement de la catégorisation des systèmes de recommandation, comprend notamment le filtrage collaboratif utilisateur-utilisateur et objet-objet.

Les défis de conception, d'approche et d'utilisation varient selon la catégorie du système, mais certains défis de base sont universels. Par exemple, l'analyse limitée du contenu pose problème pour les systèmes basés sur le contenu, tandis que la rareté des notations affecte spécifiquement les systèmes collaboratifs. Les nouveaux utilisateurs constituent un défi commun à tous les types de systèmes de recommandation.

En raison de leur utilisation généralisée, ces systèmes suscitent un grand intérêt, avec de nombreux efforts consacrés à comprendre et résoudre les problèmes qui leur sont associés. Dans ce contexte, il est important de ne pas réinventer la roue, mais plutôt de cibler les sous-défis clés et d'élaborer des solutions efficaces et réfléchies pour y faire face, compte tenu de la nature multidisciplinaire et évolutive de ces systèmes.

Les données

Les données et informations dans les systèmes de recommandation sont cruciales pour l'apprentissage et la génération de recommandations. Cependant, leur surabondance sans compréhension adéquate peut conduire à des approches erronées. Les défis incluent la diversité des formes de données, des opinions personnelles aux attributs des utilisateurs, nécessitant une alimentation adéquate des systèmes pour offrir des recommandations de qualité.

Le filtrage collaboratif, regroupant les expériences d'utilisateurs similaires, est une approche réussie mais sujette à des limitations telles que la parcimonie des données. Ce problème survient lorsque les données sont insuffisantes pour identifier les utilisateurs similaires, limitant la qualité des recommandations. Notre étude se

concentre sur le développement d'approches pour formuler des recommandations de haute qualité même en cas de données limitées.

Modèle d'apprentissage

Le cœur de tout système de recommandation est la boîte noire dans laquelle les données sont entrées pour avoir les sorties ordonnées prédites. Bien que la conception de la boîte noire soit sans aucun doute un travail ardu et nécessite une expertise compétitive dans le domaine, même le choix de la technologie appropriée pour la boîte noire est très exigeant.



Source : élaborer par nous-même

Les techniques de boîte noire dépendent en grande partie du type de recommandations souhaité dans la sortie et / ou de ses applications associées. Il existe encore de nombreux cas où plus d'une technologie est appropriée et laisse donc le concepteur dans le dilemme. Chacun de ces choix a ses avantages et ses inconvénients, ce qui en fait un choix cher au concepteur. Il est très difficile de répondre à cette question de « Comment apprendre ? ». Parmi les techniques d'apprentissage (et les algorithmes) les plus courantes pour les systèmes de recommandation de filtrage collaboratif, on peut citer la factorisation matricielle, la décomposition en valeurs singulières, KNN Voisins et factorisation matricielle non négative. Dans le même esprit, les méthodes de classification et d'apprentissage probabiliste issues du machine Learning sont très populaires pour les recommandations basées sur le contenu.

Un autre facteur important est la densité des données impliquées dans les calculs. Certaines techniques conviennent mieux aux données à haute densité qu'aux données sparse data et inversement. En plus de cela, des éléments tels que le temps de calcul, la facilité de mise en œuvre, la qualité de la recommandation et la résilience face aux pannes sont également des aspects cruciaux à prendre en compte. Ces considérations sont particulièrement pertinentes pour choisir parmi les approches les plus populaires, comme mentionné précédemment.

Confidentialité

Les systèmes de recommandation ont gagné en popularité auprès des consommateurs et des fournisseurs de services. Lorsqu'ils sont conçus efficacement, ils peuvent être des outils extrêmement puissants pour la personnalisation. Bien que, outre la conception efficace, les données d'entrée jouent un rôle important. Les détails personnels, les traits précis, les opinions exactes et les préférences des utilisateurs sont parmi les informations les plus recherchées pour faire ces recommandations personnalisées. C'est devenu une tendance des fournisseurs de services à ignorer souvent les risques et les atteintes à la vie privée liés à la collecte et au traitement des données personnelles des utilisateurs. En outre, de nombreux utilisateurs ne sont pas suffisamment conscients des menaces à la vie privée auxquelles ils pourraient être exposés.

Les deux principaux facteurs responsables de cette situation sont les suivants :

- Ne pas être pleinement conscient de toutes les données collectées.
- Sous-estimation de la puissance des données collectées et des modèles d'apprentissage.

Par exemple, un utilisateur qui utilise Amazon pour acheter un appareil photo peut ne pas savoir que son historique de navigation est en cours de sauvegarde, même s'il n'a pas apporté le produit. De même, un utilisateur écrivant un avis sur un produit acheté peut ne pas savoir que, par exemple, l'avis combiné à ses informations personnelles comme le sexe et l'âge peut révéler sa ville ou toute autre information inconnue et non déclarée la concernant, les informations de l'utilisateur souvent piraté ce qui pose le problème de menace à la vie privée du client et la confiance des systèmes de recommandations.

Les critères du choix

Bien qu'un système de recommandation soit en soi un logiciel complexe, il fait naturellement partie d'un plus grand système. Un élément de recommandation est l'une des caractéristiques d'une application globale. Ce système peut être une fonctionnalité mineure ou un principal argument de vente ; l'application peut-être déjà existante ou construite avec le recommandeur, mais, dans tous les cas, la conception d'un système de recommandation doit être intégrée à la conception de l'application.

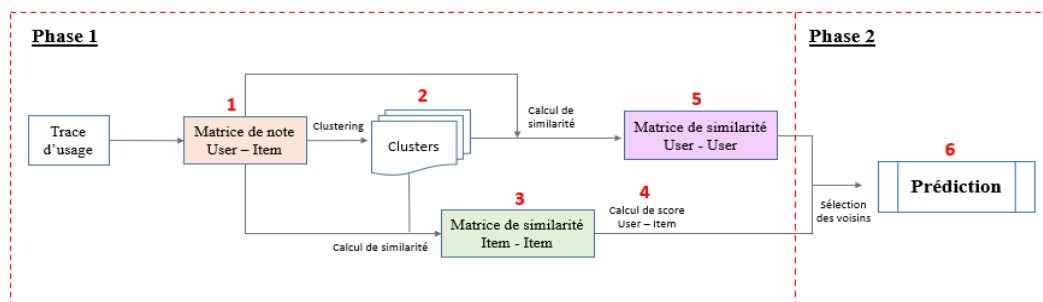
L'hébergement Cette section étudie les principaux facteurs relatifs à l'application hôte qui devraient influencer Le concepteur de recommandations suit deux axes principaux : le rôle du recommandeur et l'influence de la mise en œuvre de l'application.

Afin de comprendre le rôle du système de recommandation dans le portail, Il fallait répondre aux questions suivantes :

- Quel est le but du moteur de recommandation ?
- Quel est le type de recommandations qu'il devrait fournir ?
- Quelles mesures seront utilisées pour mesurer les performances du moteur de recommandation ?

VI. Application Potentielle Du Modèle

Architecture du système de recommandation



Source : élaborer par nous-même

Afin de concrétiser ce scénario, nous allons mettre en œuvre un moteur de recommandation de produits basé sur l'approche du filtrage collaboratif (FC). Le choix du FC s'explique par deux raisons principales :

Tout d'abord, le jeu de données ne comporte que des informations relatives aux utilisateurs et aux produits qu'ils utilisent. En conséquence, l'utilisation du filtrage basé sur le contenu (BSC) n'est pas envisageable, car cette méthode requiert des évaluations des produits ainsi que des informations détaillées sur les articles, telles que les attributs.

Deuxièmement, le filtrage collaboratif est reconnu comme l'un des moyens les plus efficaces lorsqu'il s'agit de formuler des recommandations pour des utilisateurs déjà connus.

En optant pour cette approche, notre objectif est de mettre en place un moteur de recommandation performant, adapté à la nature spécifique de notre jeu de données, tout en capitalisant sur les avantages du filtrage collaboratif pour offrir des suggestions de produits pertinentes et efficaces.

Le processus se divise en deux phases distinctes : la "**PHASE I**" correspond à la phase d'**apprentissage**, tandis que la "**PHASE II**" représente la **prédiction**.

→ **Voici les étapes détaillées:**

- 1) Exploitation de la matrice de notation "Utilisateur - Article", qui contient les évaluations estimées déduites des données historiques.
- 2) Mise en œuvre d'un regroupement (clustering) basé sur cette même matrice afin de créer des groupes d'utilisateurs partageant des caractéristiques similaires.
- 3) Calcul d'une matrice de similarité basée sur les articles pour chaque cluster, en se basant sur les résultats des deux étapes précédentes.
- 4) Utilisation de la matrice "Utilisateur - Article" pour attribuer à chaque utilisateur un score spécifique par rapport à chaque article, permettant ainsi la définition des meilleurs produits recommandés pour les utilisateurs appartenant au même cluster.
- 5) Tentative de calcul de similarité, mais cette fois-ci, en employant une similarité utilisateur-utilisateur. Ceci consiste en une comparaison entre chaque nouveau client et l'historique représenté par la matrice utilisateur-article. L'objectif est de recommander à un nouveau client des produits déjà appréciés par d'autres clients similaires.
- 6) Identification des prédictions, matérialisées sous forme de recommandations, finalisant ainsi le processus.

En suivant ces étapes, le processus vise à tirer parti des informations contenues dans la matrice utilisateur-article pour créer des recommandations pertinentes, en prenant en compte à la fois les évaluations passées et les caractéristiques similaires entre utilisateurs.

Solution

Ces choix font partie de l'environnement de la recommandation.

L'environnement de la recommandation peut être étudié à travers quatre dimensions :

- **Utilisateurs** : qui sont les utilisateurs, quels sont leurs objectifs ?
- **Données** : quelles sont les caractéristiques des données sur lesquelles sont basées les recommandations ?

- **Application** : à quelle application globale le recommandeur appartient-il ?
- **Algorithmes** : Quelles méthodes de recommandation utilisées ?

L'approche que nous allons utiliser pour notre moteur de recommandation est une approche filtrage collaboratif. La raison pour laquelle nous avons adopté cette approche réside dans l'analyse des avantages et des inconvénients des différentes versions des systèmes de recommandation

L'analyse a montré les résultats suivants :

- **Approche basée sur le contenu** : on n'a pas les données nécessaires pour utiliser cette approche.
- **Approche hybride** : demande au moins deux systèmes de recommandations, comme première phase, par conséquent, elle n'est pas une bonne solution.
- **Pour les données** : On a choisi les données démographiques (identifiant du client, âge, sexe, état matrimonial, profession, strate, l'activité ...) et l'historique des utilisateurs dans la sélection des services et produits.

Choix des technologies

→ **Technologies et outils du développement :**

Langages de programmation	Bibliothèques
	

VII. Discussion

La transformation numérique a un impact majeur sur la relation client dans le secteur de l'assurance, comparable à la révolution informatique précédente qui avait principalement remodelé les structures internes des compagnies. L'émergence des AssurTech réinvente les modèles commerciaux traditionnels en exploitant la technologie, transformant directement les interactions entre compagnies et clients. Cette évolution stimule la distribution de l'assurance à travers divers canaux, allant de la démocratisation des agences traditionnelles à des partenariats innovants avec des producteurs locaux et des collaborations avec des start-ups.

Cependant, cette transformation suscite des interrogations chez les professionnels de l'assurance concernant l'adéquation des processus traditionnels aux attentes d'une génération de digital natives, ainsi que sur la pérennité de leurs rôles face à l'évolution rapide des technologies, des données et des algorithmes (Lohrmann, 2019). Pour répondre à ces préoccupations, l'ACPR (2022) suggère que les compagnies d'assurance doivent développer les compétences en conseil de leurs employés et proposent des formations adaptées. L'exemple du directeur d'agence illustre l'impact tangible de la transformation numérique sur le métier des assureurs. La digitalisation croissante dans la vie quotidienne a considérablement modifié ce rôle. Les outils numériques permettent désormais aux directeurs d'agence de déléguer davantage de responsabilités à leurs conseillers, renforçant ainsi leur autonomie, et simplifient la gestion de l'agence grâce à des indicateurs de performance accessibles via des applications mobiles. Cette évolution est soutenue par plusieurs études montrant que l'introduction de nouvelles technologies entraîne l'émergence de nouveaux métiers et redéfinit les dimensions et les tâches dans le secteur de la bancassurance. Ainsi, le directeur d'agence, grâce au numérique, peut adopter une approche plus centrée sur le client et se concentrer de manière proactive sur les activités de prospection.

VIII. Conclusion générale

La transformation digitale est importante pour les compagnies d'assurance cherchant à rester compétitives. Les AssurTech proposent des solutions innovantes, tandis que des éléments comme l'amélioration de l'expérience client, des processus opérationnels efficaces et des modèles commerciaux adaptés restent essentiels. Il est important de distinguer l'impact du digital sur l'écosystème externe et interne des entreprises d'assurance pour évaluer les opportunités et les menaces du marché. Les smartphones et les applications mobiles ont impulsé la finance mobile, avec des implications majeures pour l'assurance.

La crise financière de 2008 a favorisé l'émergence de la Fintech, modifiant la réglementation et encourageant l'innovation. La relation client-assureur évolue vers une expérience client plus centrée. Les professionnels de l'assurance doivent développer de nouvelles compétences pour répondre aux besoins en adaptation aux nouvelles technologies. Les systèmes de recommandation basés sur le Big Data et l'apprentissage automatique personnalisent l'expérience client. Le clustering est essentiel en analyse des données pour regrouper des données similaires et segmenter les marchés. La transformation digitale est un processus évolutif, nécessitant une adaptation continue pour rester compétitif dans un paysage en mutation rapide.

IX. Références:

- [1] Deloitte. (2024). Étude sur la personnalisation des offres d'assurance. Perspectives pour le secteur de l'assurance en 2024 : Les assureurs évoluent pour s'adapter au contexte d'exploitation changeant et avoir une incidence sociétale encore plus grande.
URL : <https://www2.deloitte.com/fr/fr/pages/services-financier/articles/perspectives-pour-le-secteur-assurance-en-2024.html>
- [2] Barbara, C., Cortis, D., Perotti, R., Sammut, C., Vella, A., Barbara, C., ... Vella, A. (2017). The European Insurance Industry: A PEST Analysis. *International Journal of Financial Studies*, 5(2), 14.
<https://doi.org/10.3390/ijfs5020014>
- [3] Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, 70, 328–337.
<https://doi.org/10.1016/j.jbusres.2016.08.006>
- [4] Bittini JS, Rambaud SC, Pascual JL, Moro-Visconti R. Business models and sustainability plans in the FinTech, InsurTech, and PropTech industry: Evidence from Spain. *Sustainability*. 2022; 14(19): 12088.
- [5] Charlin, L. (2017). Intelligence artificielle : une mine d'or pour les entreprises. *Gestion*, 42(1), 76.
<https://doi.org/10.3917/riges.421.0076>
- [6] Su, Xiaoyuan, & Khoshgoftaar, Taghi. (2009). A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*, 2009. DOI: 10.1155/2009/421425
- [7] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35, 137–144.
<https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- [8] YILDIRIM İ, Şahin EE. Insurance Technologies (Insurtech): Blockchain and Its Possible Impact on Turkish Insurance Sector. *Journal of International Management Educational and Economics Perspectives*. 2018; 6(3): 13-22.
- [9] Halima EH, Yassine T. Insurtech & blockchain: implementation of technology in insurance operations and its environmental impact. In *IOP Conference Series: Earth and Environmental Science*; 2022: IOP Publishing. p. 012010.
- [10] Cappiello A, Cappiello A. Digital disruption and insurtech start-ups: Risks and challenges. *Technology and the Insurance Industry: Re-configuring the Competitive Landscape*. 2018;2020: 29-50.
- [11] Kalika, M., & Kefi, H. (2004). Evaluation des systèmes d'information: une perspective organisationnelle. Post-Print. Retrieved from <https://ideas.repec.org/p/hal/journal/hal00155610.html>
- [12] Kalika, M., & Kefi, H. (2004). Evaluation des systèmes d'information: une perspective organisationnelle. Post-Print. Retrieved from <https://hal.science/hal-00155610>
- [13] Bouveyron C, Celeux G, Murphy TB, & Raftery AE 2019, *Model-Based Clustering and Classification for Data Science: With Applications in R* (Cambridge: Cambridge Univ. Press
- [13] ACPR, (2022). « La transformation numérique dans le secteur français de l'assurance », Analyses et synthèses, n° 132, Paris, France.

Webographies :

- La transformation digitale dans l'assurance : enjeux et perspectives 2024. *Journal la tribune now*. URL : <https://www.latribune.fr/supplement/la-tribune-now/la-transformation-digitale-dans-l-assurance-enjeux-et-perspectives-2024-994891.html> (consulté en 23/02/2024)
- Huret, Louise. Les systèmes de recommandation et leur impact sur le phénomène de surconsommation. Louvain School of Management, Université catholique de Louvain, 2020. Prom. : Fouss, François.
<http://hdl.handle.net/2078.1/thesis:25678> (consulté en 23/02/2024)

Ouvrages :

- The InsurTech Book, the Insurance Technology Handbook for Investors, Entrepreneurs and FinTech Visionaries, Edited by Sabine L. B. VanderLinden, Shân M. Millie, Nicole Anderson, 2018.
- The Current InsurTech Landscape: Business Models and Disruptive Potential, Alexander Braun, Florian Schreiber, 2017