

2023 - 2024
PROJET DE FIN D'ÉTUDES

DIPLÔME NATIONAL D'INGÉNIEUR

SPECIALITÉ: INFORMATIQUE

**Système de recommandation de
produits en cristal 3D**

Réalisé par: YOSRI BEN SALEM

Encadré par:

Encadrant ESPRIM: Mme. FEHIMA ACHOUR

Encadrant Entreprise: M. MOHAMED BOURAOUI

Remerciements

Je tiens à exprimer mes sincères remerciements à **mes parents** pour leur soutien inconditionnel tout au long de mon parcours académique. Leur amour, leurs encouragements et leur dévouement ont été une source d'inspiration et de motivation essentielle pour moi.

Je souhaite également exprimer ma gratitude envers la **société Bouraoui Group** et son co-fondateur, **Mr. Mohamed Bouraoui**, pour leur généreux soutien Technique et leur confiance en mes capacités. Leur engagement en faveur de l'éducation et du développement des jeunes talents est exemplaire, et je suis honoré d'avoir été bénéficiaire de leur soutien.

En outre, je tiens à remercier chaleureusement l'ensemble des enseignants de l'**ESPRIM** pour leur expertise, leurs conseils précieux et leur dévouement à notre apprentissage. Leur passion pour l'enseignement et leur engagement à nous fournir une éducation de qualité ont grandement contribué à ma formation et à ma croissance personnelle et professionnelle.

Yosri Ben Salem

table des matières

Introduction générale	1
1 Cadre et contexte du projet	3
1.1 Introduction	3
1.2 Contexte et étude de l'existant	3
1.2.1 Problématique	3
1.2.2 Etude de l'existant	4
1.2.3 Solution proposée et objectifs	7
1.3 Présentation Générale de l'entreprise	8
1.4 Langage de modélisation et méthodologie de conception	9
1.4.1 Méthodes Agiles	9
1.4.2 CRISP-DM	9
1.4.3 SCRUM	11
1.4.4 Langage de modélisation UML	12
1.5 Etat de l'art	12
1.5.1 Intelligence Artificielle	12
1.5.2 Machine Learning	13
1.5.3 Deep Learning	13
1.6 Conclusion	14
2 Compréhension du métier	15
2.1 Introduction	15
2.2 Description du domaine des produits 3D Crystal	15
2.3 Analyse des besoins	16
2.3.1 Identifications des acteurs	16
2.3.2 Besoins fonctionnels	17
2.3.3 Besoins non fonctionnels	18
2.4 Objectifs et Contraintes du projet	18
2.4.1 Objectifs du système de recommandation :	19
2.4.2 Contraintes du projet	19
2.5 Backlog produit	20
2.6 Planification des sprints	22
2.7 Conclusion	23

3	Exploration des données	24
3.1	Introduction	24
3.2	Présentation des sources de données	24
3.2.1	Base des produits 3D Crystal	24
3.2.2	Ensemble de questions-réponses préétablies	25
3.2.3	Données d'entrée utilisateur pour le NLP	26
3.3	Analyse des données	27
3.3.1	Analyse du fichier CSV des produits 3D Crystal	27
3.3.2	Étude des questions-réponses	28
3.3.3	Préparation des données textuelles pour la recommandation	29
3.3.4	Étapes du prétraitement	30
3.4	Conclusion	30
4	Modélisation et développement	31
4.1	Introduction	31
4.2	Modélisation du recommandation par questions-réponses	31
4.2.1	Méthodes utilisées	31
4.2.2	Prétraitement des données	34
4.2.3	Transformation en sacs de mots	35
4.2.4	Application aux questions-réponses	35
4.3	Modélisation du recommandation par NLP	36
4.3.1	Principe du NLP	36
4.3.2	Principaux domaines d'application du NLP	37
4.3.3	Les principaux algorithmes NLP	38
4.3.4	Classifieur zéro-shot	40
4.4	Conclusion	41
5	Evaluation et Déploiement	42
5.1	Introduction	42
5.2	Critères d'évaluation théorique	42
5.3	Cycle de vie du projet	43
5.4	Méthodologie d'évaluation	44
5.4.1	Prédiction au niveau des phrases	44
5.4.2	Agrégation des prédictions au niveau des textes	44
5.4.3	Calcul de l'exactitude	45
5.5	Déploiement	45
5.5.1	Technologies utilisées	45
5.5.2	Interfaces	47
5.5.3	Hébergement	49
5.6	Conclusion	50

TABLE DES MATIÈRES

Conclusion et perspectives	51
bibliographie	53
A	54
B	55
C	56

Liste des Figures

1.1	Artpix 3D Logo	5
1.2	Crystal Clear Memories Logo	5
1.3	Logo 3dcrystal.com	6
1.4	Logo Société Bouraoui Group	8
1.5	Les phases principales du Méthodologie CRISP-DM	11
1.6	Illustration de la méthodologie SCRUM	11
3.1	Application Shopify : Ablestar Bulk Product Editor	25
4.1	Processus de la méthode Similarité cosinus	32
4.2	Simulation de l'algorithme KNN	33
4.3	principe du modèle TF-IDF	35
4.4	Domaines d'application du NLP	37
4.5	Zero-Shot Learning	40
5.1	Cycle de vie du projet	43
5.2	Logo Python	45
5.3	Logo Flask	46
5.4	Logo HTML	46
5.5	Logo CSS	46
5.6	Logo Javascript	47
5.7	Page d'accueil	47
5.8	Page de recommandation basée sur le traitement des ques- tions/réponses	48
5.9	Page de recommandation basée sur le traitement du text NLP	49
5.10	Logo Heroku	50

Liste des tableaux

Backlog produit	21
Planification des sprints	23
Tableau des questions-réponses	26
Tableau comparatif des algorithmes NLP	40

Introduction générale

Le secteur des produits 3D Crystal connaît une popularité croissante, offrant aux consommateurs la possibilité de créer des objets personnalisés et uniques. Cependant, avec la multitude d'options disponibles, les utilisateurs peuvent se retrouver confrontés à un défi de choix. Dans cette optique, nous proposons la mise en place d'un système de recommandation de produits 3D Crystal avec un volet de questions/réponses et de traitement du langage naturel (NLP). Ce système vise à simplifier le processus de sélection des produits en offrant aux utilisateurs des recommandations personnalisées basées sur leurs préférences et leurs besoins spécifiques.

L'objectif principal de ce projet est de développer un système de recommandation intelligent et efficace, capable de comprendre les préférences des utilisateurs à partir de leurs réponses à des questions spécifiques et d'utiliser des techniques de traitement du langage naturel pour interpréter et analyser ces réponses. Grâce à ce système, les utilisateurs pourront bénéficier d'une expérience de sélection de produits plus personnalisée et pertinente, tout en réduisant le temps et les efforts nécessaires pour trouver le produit idéal.

Pour atteindre cet objectif, nous adoptons une approche basée sur la méthodologie CRISP-DM (Cross-Industry Standard Process for Data Mining), qui nous permettra de gérer efficacement les différentes étapes du projet, de la compréhension du problème à la modélisation et à l'évaluation du système de recommandation.

Ce rapport présente une analyse détaillée de notre démarche, en mettant l'accent sur les différentes étapes du processus CRISP-DM adaptées à notre projet. Nous commencerons par une étape de compréhension approfondie du problème, en identifiant les besoins des utilisateurs et en décrivant les données disponibles, y compris les questions/réponses et les données NLP. Ensuite, nous explorerons ces données, les prétraiterons et les analyserons pour mieux comprendre leur structure et leurs caractéristiques.

La section suivante sera consacrée à la modélisation, où nous détaillerons les techniques utilisées pour construire notre système de recommandation, en mettant l'accent sur les approches de recommandation collaborative, de filtrage basé sur le contenu et d'apprentissage automatique. Nous discuterons des choix de modèles et des algorithmes spécifiques qui sous-tendent notre système.

L'évaluation du système de recommandation joue un rôle crucial dans notre projet, et nous consacrerons donc une section spécifique à cette étape. Nous présenterons les méthodes d'évaluation utilisées, les mesures de performance

employées, ainsi que les résultats obtenus et une comparaison des performances des différents modèles.

Enfin, nous aborderons la question du déploiement du système de recommandation, en décrivant les étapes et les considérations nécessaires pour intégrer le système dans une application ou une plateforme existante. Nous conclurons ce rapport en récapitulant les résultats obtenus, en discutant des limites et des défis rencontrés, ainsi qu'en proposant des pistes d'amélioration pour notre système de recommandation.

Cadre et contexte du projet

1.1 Introduction

Dans ce premier chapitre, on décrira les parties qui permettent la définition du champ d'étude et le planning prévisionnel. Ensuite, une étude de l'existant sera nécessaire afin de dégager les critiques du système actuel. Enfin, on citera les objectifs à atteindre ainsi que les avantages de l'application.

1.2 Contexte et étude de l'existant

1.2.1 Problématique

La problématique du projet consiste à améliorer le taux de conversion d'un site proposant plus de 1000 produits, notamment des créations en cristal 3D gravées, en utilisant les technologies du machine learning. Le site fait face à des difficultés pour convertir les visiteurs en clients, ce qui limite sa croissance et sa rentabilité. L'objectif est donc d'intégrer des techniques de machine learning afin d'optimiser l'expérience utilisateur en fournissant des recommandations de produits pertinentes et personnalisées. En exploitant les données existantes sur les comportements d'achat des utilisateurs et en utilisant des algorithmes de machine learning, il sera possible de comprendre les préférences individuelles des clients, d'analyser les tendances du marché et de proposer des recommandations précises et ciblées. Ces recommandations, basées sur des modèles prédictifs, permettront de susciter l'intérêt des visiteurs en leur présentant des créations en cristal 3D gravées qui correspondent parfaitement à leurs goûts et à leurs aspirations. Cette approche personnalisée facilitera leur processus de prise de décision et augmentera ainsi le nombre de conversions.

Pour répondre à cette problématique, plusieurs défis doivent être relevés :

1. Collecte et analyse des données : L'un des défis majeurs sera de collecter et d'analyser les données sur les comportements d'achat des utilisateurs.

Cela implique la mise en place de mécanismes de suivi et de collecte de données pertinents, ainsi que le développement d'algorithmes de traitement de données efficaces pour extraire les informations nécessaires à la création de recommandations personnalisées.

2. Modélisation des préférences des utilisateurs : une étape clé pour être capable de formuler des recommandations personnalisées pertinentes. Cela nécessite la mise en place de modèles d'apprentissage automatique (machine learning) capables d'analyser les données collectées sur les utilisateurs et d'en déduire leurs centres d'intérêt récurrents.
3. Développement d'algorithmes de recommandation efficaces : La création d'algorithmes de recommandation précis et performants constitue un autre défi. Il faudra explorer différentes approches, telles que les filtres collaboratifs, les méthodes basées sur le contenu ou les techniques de deep learning, pour trouver la méthode la plus adaptée aux caractéristiques des produits en cristal 3D gravés et aux spécificités du marché.
4. Intégration des recommandations dans l'expérience utilisateur : Une fois les recommandations générées, il sera crucial de les intégrer de manière transparente dans l'expérience utilisateur du site. Cela nécessitera une réflexion sur le design de l'interface, l'emplacement des recommandations et la manière de les présenter de manière attrayante, afin d'inciter les utilisateurs à explorer les produits suggérés et à passer à l'achat.
5. Évaluation et ajustement continu : Enfin, un défi constant sera de mesurer l'efficacité des recommandations et de les ajuster en fonction des retours des utilisateurs et des performances observées. Une évaluation régulière des résultats obtenus permettra d'identifier les améliorations à apporter et d'optimiser en permanence les algorithmes de recommandation pour maximiser le taux de conversion.

1.2.2 Etude de l'existant

1.2.2.1 Analyse de l'existant

En termes d'analyse de l'existant, les exemples mentionnées ci-dessous montrent trois des concurrents en ligne qui offre des services similaires :

Artpix 3d



FIGURE 1.1 – Artpix 3D Logo

ArtPix 3D offre une large sélection de cadeaux personnalisés en cristal gravé en 3D, avec la possibilité de prévisualiser les modèles avant l'achat. Leur engagement envers la satisfaction client, la livraison gratuite et la garantie de retour de 14 jours sont des points positifs. Cependant, en ce qui concerne la présentation des produits, le site manque peut-être d'un design plus moderne et attrayant. Les produits sont listés de manière assez standard, ce qui peut rendre la recherche un peu fastidieuse pour les clients qui souhaitent trouver rapidement ce qu'ils cherchent.

Crystal Clear Memories



FIGURE 1.2 – Crystal Clear Memories Logo

Crystal Clear Memories est un site web spécialisé dans la vente de cristaux gravés en 3D, offrant aux clients la possibilité de personnaliser leurs propres photos sur ces cristaux. Cependant, le site ne semble pas proposer un système de recommandation de produits 3D gravés en cristal spécifique. Le site offre

une large gamme de produits, avec une sélection de cadeaux telles que des tours, des briques, des cubes, des cœurs et des modèles de prestige. Cependant, il manque un mécanisme qui recommanderait spécifiquement des produits 3D gravés en cristal en fonction des préférences et des besoins des clients.

3DCrystal.com



FIGURE 1.3 – Logo 3dcrystal.com

3DCrystal.com présente plusieurs points forts dans son activité de vente de cadeaux en cristal gravés en 3D. Il offre une large gamme de produits personnalisables, permettant aux clients de transformer leurs précieux souvenirs en pièces d'art uniques. La variété de formes et de tailles disponibles, comme les rectangles, les cœurs et les modèles de prestige, offre aux clients un choix diversifié. Cependant, certaines critiques peuvent être formulées à l'égard de 3DCrystal.com. Certains clients pourraient souhaiter un système de recommandation plus spécifique pour les produits gravés en cristal, basé sur leurs préférences et leurs besoins. De plus, l'absence d'informations détaillées sur les matériaux utilisés et les processus de fabrication peut laisser certains clients en quête de transparence.

1.2.2.2 Critique de l'existant

En analysant l'existant dans le domaine des cadeaux personnalisés en cristal gravé en 3D, il est possible de relever certains éléments. Le site examiné propose une large sélection de produits avec la possibilité de personnaliser les photos sur les cristaux. Cependant, il manque un système de recommandation spécifique aux produits 3D gravés en cristal. Cette absence limite la capacité du site à proposer des suggestions pertinentes et adaptées aux préférences des clients.

Un système de recommandation efficace est crucial pour aider les clients à naviguer parmi les nombreuses options disponibles et à trouver rapidement les produits qui correspondent le mieux à leurs préférences et à leurs besoins. En l'absence d'un tel système, le site se limite à simplement présenter une liste

standard de produits, sans prendre en compte les préférences individuelles des clients. Cela peut rendre la recherche fastidieuse et décourageante, surtout pour les clients qui ne savent pas exactement ce qu'ils cherchent.

1.2.3 Solution proposée et objectifs

Pour combler le manque d'un système de recommandation spécifique aux produits 3D gravés en cristal, une solution intéressante et efficace serait de combiner un chatbot interactif basé sur des questions-réponses avec des algorithmes de recommandation intelligents. Lorsqu'un client visite le site, le chatbot se présente et engage une conversation avec lui. Il poserait des questions ciblées pour comprendre les préférences de l'utilisateur en termes de centres d'intérêts, de cible, d'émotions, de style etc. Ces questions seraient présentées de manière conviviale et engageante, permettant aux clients de répondre facilement et de manière ludique.

En fonction des réponses fournies par l'utilisateur, le chatbot utiliserait des algorithmes de recommandation pour générer des suggestions personnalisées. Les algorithmes pourraient prendre en compte les préférences déclarées par l'utilisateur ainsi qu'il peut prédire les labels des préférences pour les cadeaux A proposer à partir d'analyse du langage naturel du texte saisi par l'utilisateur. Cela permettrait de proposer des recommandations pertinentes et adaptées aux goûts de chaque client.

Ci-dessous, vous trouverez une liste d'objectifs relatifs à la mise en place :

- Collecter les préférences des utilisateurs : L'objectif principal est de collecter de manière précise et complète les préférences des utilisateurs en termes D'occasion ,intérêt, émotions ,audience, etc. Cela permettra de comprendre leurs goûts et leurs besoins spécifiques.
- Générer des recommandations personnalisées : L'objectif est d'utiliser les réponses fournies par les utilisateurs pour générer des suggestions de produits personnalisés. Les recommandations doivent être pertinentes et adaptées aux préférences de chaque client.
- Améliorer l'expérience d'achat : L'objectif est d'améliorer l'expérience d'achat des clients en leur fournissant des recommandations personnalisées qui correspondent à leurs goûts et à leurs besoins. Cela facilitera leur processus de recherche et de prise de décision, en leur proposant des produits pertinents et attrayants.
- Encourager la découverte de nouveaux produits : L'objectif est de stimuler l'exploration des options disponibles par les clients en leur proposant

des produits qu'ils n'auraient peut-être pas considérés initialement. Les recommandations peuvent inclure des produits similaires ou complémentaires qui pourraient susciter l'intérêt des clients et les inciter à découvrir de nouvelles possibilités.

- Optimiser les recommandations au fil du temps : L'objectif est d'améliorer la précision des recommandations en utilisant des algorithmes d'apprentissage automatique qui apprennent des interactions précédentes avec les clients. Le système doit être capable de s'adapter et de fournir des recommandations de plus en plus précises à mesure qu'il collecte davantage de données et de feedback des utilisateurs.
- Augmenter les ventes et la satisfaction client : L'objectif ultime est d'augmenter les ventes en proposant des produits pertinents et attrayants aux clients. En fournissant des recommandations personnalisées et en améliorant l'expérience d'achat, le site peut augmenter la satisfaction globale des clients, favoriser la fidélité et encourager les achats répétés.

1.3 Présentation Générale de l'entreprise



FIGURE 1.4 – Logo Société Bouraoui Group

Société Bouraoui Group est une entreprise dynamique et innovante dédiée à fournir des services de haute qualité dans divers domaines. Avec une forte focalisation sur l'excellence, nous nous efforçons de répondre aux besoins uniques et aux attentes de nos clients. Notre expertise s'étend à travers diverses industries, y compris l'architecture, le design et la technologie.

Au sein du Société Bouraoui Group, nous nous engageons à offrir des solutions de pointe et à promouvoir les avancées créatives. Notre équipe est composée de professionnels hautement qualifiés avec une vaste expérience dans

leurs domaines respectifs. Nous sommes fiers de rester à la pointe des tendances de l'industrie et d'utiliser les dernières technologies pour fournir des résultats exceptionnels.

En tant qu'organisation centrée sur le client, nous accordons la priorité à la construction de relations solides et à la compréhension des exigences spécifiques de chaque projet. Que ce soit pour la **visualisation architecturale**, le **rendu 3D** ou d'autres services de **photographie**, nous abordons chaque tâche avec une attention méticuleuse aux détails et une passion pour dépasser les attentes des clients.

1.4 Langage de modélisation et méthodologie de conception

Avant la réalisation de tout projet informatique, il est nécessaire de choisir une méthodologie de travail et un processus de suivi afin d'aboutir à la fin à un logiciel fiable. Cette méthodologie présente un procédé qui a pour objectif de permettre de formaliser les étapes préliminaires du développement d'un système afin de rendre ce développement plus fidèle aux besoins du client. Il existe plusieurs méthodes dans ce qui suit, nous nous intéressons aux méthodes agiles vues leurs apports pour des projets ou les besoins évoluent.

1.4.1 Méthodes Agiles

«Les méthodes agiles partent du principe que spécifier et planifier dans les détails l'intégralité d'un produit avant de le développer (approche prédictive) est Contre-productif.»[1]

Cette méthode s'agit de la réitération de cycles courts dans le temps en divisant le projet en de multiples mini-projets et les prioriser selon les besoins.

1.4.2 CRISP-DM

CRISP-DM (Cross-Industry Standard Process for Data Mining) est un modèle de processus standardisé largement utilisé dans le domaine de l'exploration de données et de l'analyse prédictive. Il fournit une approche structurée pour guider les équipes dans la réalisation de projets d'exploration de données, de la compréhension des besoins commerciaux à la mise en œuvre des solutions.

CRISP-DM se compose de six phases principales, qui sont les suivantes :

- **Compréhension du domaine** : Cette phase consiste à définir les objectifs commerciaux et les besoins en matière de données, ainsi qu'à comprendre le domaine d'application dans lequel le projet d'exploration de données est mené.
- **Compréhension des données** : Ici, on rassemble les données nécessaires au projet, on les explore, on les nettoie et on les prépare pour l'analyse.
- **Préparation des données** : Dans cette phase, les données sont transformées et préparées pour l'analyse ultérieure. Cela peut inclure des activités telles que la sélection des variables, la création de nouvelles caractéristiques, la normalisation des données, etc.
- **Modélisation** : Cette phase implique le développement de modèles prédictifs ou descriptifs en utilisant différentes techniques d'exploration de données, telles que la régression, les arbres de décision, les réseaux neuronaux, etc.
- **Évaluation** : Une fois que les modèles ont été construits, ils sont évalués en utilisant des critères appropriés. Cela peut inclure des mesures de précision, de rappel, de F-mesure, de matrice de confusion, etc. L'objectif est de sélectionner le modèle le plus performant.
- **Déploiement** : Enfin, le modèle sélectionné est mis en œuvre dans l'environnement de production. Cela peut impliquer le développement d'applications logicielles, l'intégration dans les systèmes existants ou la création de rapports et de visualisations pour une utilisation pratique.

La figure ci-dessous présente les phases principales du méthodologie CRISP-DM sous forme d'un schéma :

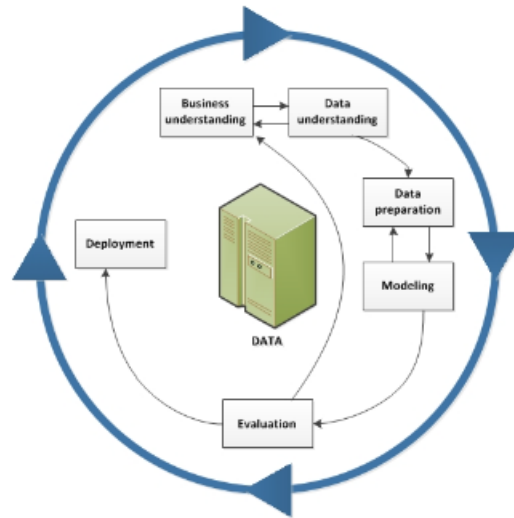


FIGURE 1.5 – Les phases principales du Méthodologie CRISP-DM

1.4.3 SCRUM

Cette méthode repose sur la gestion du temps et de l'intensité des efforts. Elle est conçue pour répondre aux besoins de la science des données en offrant des résultats rapides, rentables et prêts à être mis en production.

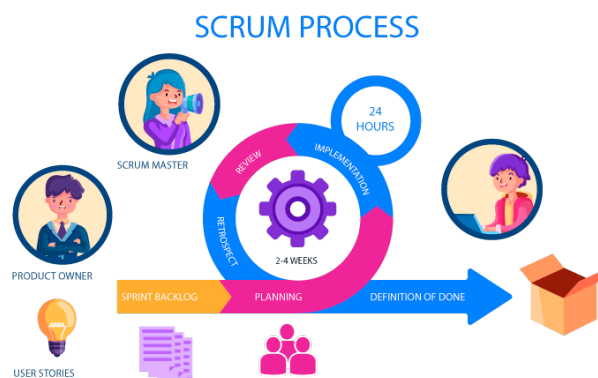


FIGURE 1.6 – Illustration de la méthodologie SCRUM

l'infrastructure Scrum ajoute cinq valeurs visant à rendre le travail efficace et la collaboration entre le acteurs en présence fluide et en ligne avec les objectifs à atteindre :

La focalisation. L'équipe projet doit être pleinement focalisée sur le développement à réaliser.

L'ouverture. L'équipe projet comme le management doivent être ouverts sur le mode de fonction de Scrum, en particulier sur la communication inter-personnel pour avancer et résoudre les problèmes ensemble.

Le respect. L'ensemble des parties prenantes (membres de l'équipe projet, management et clients) doivent faire preuve de respect mutuel.

Le courage. L'équipe projet doit faire preuve de courage pour relever les défis auxquels elle aura à faire face en toute autonomie.

L'engagement. C'est une valeur contribuant également à la réussite de la démarche. Les membres de l'équipe Scrum doivent s'engager personnellement à atteindre les objectifs de chaque sprint.

1.4.4 Langage de modélisation UML

Une dizaine d'années après le début de son utilisation dans le cadre de projets de développement orienté objet, UML[2] s'est imposé comme standard. Ce langage, dont le logo est présenté dans la figure 1.3, est né de la fusion de plusieurs méthodes existantes auparavant et est devenu désormais la référence en matière de modélisation objet. La modélisation objet consiste à créer une représentation informatique des éléments du monde réel auxquels on s'intéresse, sans se préoccuper de l'implémentation, ce qui signifie indépendamment d'un langage de programmation. Il s'agit donc de déterminer les objets présents et d'isoler leurs données et les fonctions qui les utilisent. Donc, après le choix de la méthodologie, on a opté UML comme un langage de modélisation qui est utilisé dans tous les projets logiciels comportant un ensemble des diagrammes, il permet de fournir une représentation informatique d'un ensemble d'objets et de problèmes standards du monde réel.

1.5 Etat de l'art

1.5.1 Intelligence Artificielle

L'intelligence artificielle, communément appelée IA, est une branche de l'informatique qui consiste à programmer la machine pour qu'elle devienne intelligente. Le cerveau humain est l'intuition et la motivation de l'IA, car les humains sont l'espèce la plus intelligente de la planète Terre et la sauce secrète derrière tout cela est notre puissant cerveau. Le Machine Learning (ML) et le Deep Learning (DL) sont deux sous-domaines de cette science. Bien que ces termes soient liés, il existe des distinctions entre eux.

1.5.2 Machine Learning

Le ML est un sous-ensemble de l'IA qui utilise des algorithmes d'apprentissage statistique pour créer des systèmes capables d'apprendre et de s'améliorer automatiquement à partir d'expériences sans être explicitement programmés. De manière générale, selon les données utilisées pour alimenter les algorithmes, le processus d'apprentissage de ces systèmes peut être soit :

- Apprentissage supervisé : Les résultats souhaités (des données étiquetées) sont inclus dans les données d'entraînement de la machine. Le vecteur d'entrée est présenté au réseau, la sortie est calculée et comparée au résultat souhaité. Ensuite, les poids sont ajustés pour réduire l'écart entre eux.
- Apprentissage non-supervisé : C'est un apprentissage qui n'a pas des classes préétablis. Afin de créer des groupes homogènes à partir de l'observation, il identifie des motifs communs. Le réseau s'auto-organise pour maximiser les coûts.
- Apprentissage par renforcement : consiste à apprendre les actions à prendre, à partir d'expériences, de façon à optimiser une récompense quantitative au cours du temps. Elle est le type d'apprentissage le plus ambitieux. On prend également en considération les modèles de réapprentissage automatique : support vecteur machine, arbre de décision, modèle linéaire...

1.5.3 Deep Learning

Le Deep Learning est un domaine de Machine Learning. Dans lequel au lieu de développer un des modèles que nous venons de mentionner, nous développons à la place des réseaux de neurones artificiels. Il repose sur les mêmes fondations que celles du Machine Learning : fournir des données et utiliser un algorithme d'optimisation pour ajuster le modèle à ces données. Le Deep Learning n'est pas une simple fonction mais plutôt un réseau de fonction connectées les unes aux autres. Et plus ces réseaux sont profonds, c'est-à-dire ils contiennent plus de fonctions à l'intérieur, plus la machine est capable d'apprendre à réaliser des tâches complexes comme connaître des objets, Classifier des textes, traiter les sentiments, identifier des personnes sur une photo, conduire une voiture...

Une méthode particulièrement intéressante dans ce domaine est la classification zero-shot (ou classification à zéro tir), qui repose sur le principe de généralisation à partir de connaissances préalables.

Dans la classification zero-shot, contrairement aux approches traditionnelles où les modèles sont entraînés sur des données étiquetées spécifiques à chaque classe, on cherche à classer des échantillons dans des catégories pour lesquelles le modèle n'a pas été directement entraîné. Cela est rendu possible grâce à la représentation sémantique des classes et à l'utilisation de modèles de langage.

Les réseaux de neurones profonds sont particulièrement adaptés à la classification zero-shot en raison de leur capacité à extraire des caractéristiques complexes et à apprendre des représentations hiérarchiques de données. En utilisant des techniques telles que l'apprentissage de représentations profondes et l'utilisation de réseaux neuronaux pré-entraînés sur de grandes quantités de données, il est possible de capturer les relations sémantiques entre les différentes classes et d'effectuer des classifications zero-shot précises.

1.6 Conclusion

Ce chapitre introductif nous a permis de mettre notre projet dans son cadre général. En premier lieu, nous avons présenté l'organisme d'accueil. Par la suite nous avons passé à l'analyse détaillée de l'existant puis nous avons proposé notre solution tout en précisant les objectifs visés. Enfin nous avons présenté la méthodologie de conception adoptée pour le développement de notre système et l'état de l'art. Le chapitre suivant sera consacré à l'étude des besoins fonctionnels et non fonctionnels et l'identification des données disponibles.

Compréhension du métier

2.1 Introduction

Ce chapitre présente la phase de départ de notre application. Tout d'abord, nous décrivons le domaine d'activité de l'entreprise et ses produits afin de cadrer le périmètre fonctionnel du futur système. Puis, nous listons les besoins fonctionnels et non fonctionnels de notre système ainsi que l'identification des acteurs. Ensuite, nous identifions les principaux objectifs métier et contraintes auxquels devra répondre le système de recommandation pour assurer sa réussite. Enfin, nous clôturons ce chapitre par le backlog du produit la planification des sprints.

2.2 Description du domaine des produits 3D Crystal

Les entreprises des produits "3D Crystal" sont spécialisée dans la conception et la fabrication de produits en résine époxy décorés de motifs et figures en 3D. Leurs produits incluent :

- Objets de décoration domestique (figurines, bougeoirs, porte-clés etc.)
- Badges et insignes pour événements et conventions
- Accessoires et bijoux
- Petits meubles et luminaires

Les produits sont personnalisés sur demande grâce à l'impression 3D. Les clients peuvent choisir parmi une bibliothèque de motifs prédéfinis ou soumettre leur propre design.

La personnalisation rend chaque pièce unique. Cependant, cela complique aussi la tâche marketing pour suggérer des produits aux clients. Actuellement, les recommandations se font de manière aléatoire sur le site web.

2.3 Analyse des besoins

Dans cette section, nous allons présenter les utilisateurs de notre application tout en identifiant les besoins fonctionnels et non fonctionnels.

2.3.1 Identifications des acteurs

En génie logiciel, un acteur est toute entité qui définit le rôle d'un utilisateur ou un autre système qui attend un ou plusieurs services offerts par une application.

Dans la suite, nous allons dénombrer les acteurs interagissant avec notre application :

— Acteurs internes

Visiteur : Le visiteur du store représente l'utilisateur final qui se connecte à la plateforme avec l'intention d'être un client potentiel. Son objectif principal est d'explorer les produits disponibles, d'interagir avec l'interface utilisateur et de recevoir des recommandations personnalisées en fonction de ses préférences. Le visiteur est encouragé à visiter la page des recommandations, à répondre à des questions simples et ainsi découvrir les deux produits qui correspondent le mieux à ses attentes.

— Acteurs externes

Administrateur : L'administrateur est responsable de la gestion et de la mise à jour de la base de produits sur la plateforme. Cela comprend l'ajout de nouveaux produits, la modification des descriptions, des images et des prix, ainsi que la suppression des produits obsolètes. L'administrateur doit s'assurer que la base de produits est à jour, précise et pertinente pour garantir des recommandations précises et de qualité aux visiteurs du store.

Développeur : Le développeur est chargé de la mise à jour et de l'amélioration de l'algorithme du système de recommandation. Son rôle consiste à analyser les données collectées auprès des visiteurs du store, à identifier des modèles et des tendances, et à mettre à jour l'algorithme pour fournir des recommandations plus précises et pertinentes. Le développeur travaille en étroite collaboration avec l'administrateur pour obtenir les données nécessaires à l'amélioration de l'algorithme. Il met également en place des tests et des évaluations pour mesurer l'efficacité

de l'algorithme et effectue les ajustements nécessaires pour optimiser les performances.

2.3.2 Besoins fonctionnels

Cette partie décrit les exigences que le système doit satisfaire d'une façon Informelle. Les fonctionnalités que nous proposons dans notre application sont les suivantes :

1. Intégration des données provenant de Shopify

- Le système doit être capable de récupérer les données des produits disponibles à partir de Shopify, y compris les descriptions, les images, les prix, etc.
- Il doit pouvoir synchroniser régulièrement les informations avec Shopify pour maintenir la base de produits à jour.

2. Gestion des questions pour analyser les préférences de l'utilisateur

- Le système doit permettre la création et la gestion des questions posées aux utilisateurs pour comprendre leurs préférences.
- Le système doit être capable de stocker les réponses des utilisateurs pour les utiliser dans l'algorithme de recommandation.

3. Développement de l'algorithme de recommandation basé sur les réponses aux questions :

- Le système doit mettre en place un algorithme qui analyse les réponses aux questions pour comprendre les préférences de l'utilisateur.
- Il doit utiliser ces informations pour générer des recommandations personnalisées, en mettant l'accent sur les produits correspondant aux préférences exprimées.

4. Affichage des recommandations personnalisées :

- Le système doit fournir une interface utilisateur conviviale pour afficher les recommandations personnalisées aux utilisateurs.
- Les recommandations doivent être présentées de manière attrayante, avec des informations claires sur les produits recommandés et leur pertinence par rapport aux préférences de l'utilisateur.
- L'interface utilisateur doit permettre aux utilisateurs d'interagir avec les recommandations, par exemple en ajoutant des produits au panier ou en les sauvegardant pour une consultation ultérieure.

5. Évaluation et ajustement de l'algorithme de recommandation :

- Le système doit inclure des mécanismes pour évaluer l'efficacité et la pertinence des recommandations fournies.
- Il doit permettre aux administrateurs ou aux développeurs d'analyser les résultats et d'apporter des ajustements à l'algorithme pour améliorer les performances.
- L'évaluation peut inclure des mesures telles que le taux de conversion, la satisfaction des utilisateurs ou d'autres indicateurs pertinents.

2.3.3 Besoins non fonctionnels

Performance

- Le système de recommandation doit être rapide et réactif, offrant des recommandations en temps réel pour une expérience utilisateur fluide.
- Le temps de chargement des pages et des recommandations ne doit pas dépasser quelques secondes pour éviter toute frustration de l'utilisateur.

Scalabilité

- Le système doit être capable de gérer un grand nombre d'utilisateurs simultanément, sans compromettre les performances.

Sécurité

- Le système doit garantir la confidentialité des données des utilisateurs, en mettant en place des mécanismes de protection des informations personnelles.

Convivialité

- L'interface utilisateur doit être conviviale et intuitive, permettant aux visiteurs du store de naviguer facilement et de comprendre les recommandations fournies.

2.4 Objectifs et Contraintes du projet

Les objectifs décrivent de manière générale ce que le système de recommandation devra permettre de réaliser. Ainsi, Les contraintes identifient les limitations et difficultés susceptibles d'impacter le développement du projet.

2.4.1 Objectifs du système de recommandation :

Afin de répondre aux enjeux métier et aux attentes de l'entreprise, le système devra remplir les objectifs suivants :

- Améliorer l'expérience utilisateur en proposant des produits personnalisés répondant à ses centres d'intérêt tels que déclarés lors de la réponse aux questions de recommandations ou du description de son intérêt
- Booster le taux de conversion en orientant davantage les visiteurs vers des produits susceptibles de les intéresser
- Soutenir la prise de décision clients en leur présentant une sélection plus restreinte de 3 à 5 produits seulement, adaptée à leur profil. L'objectif est de ne pas les submerger sous les choix.

2.4.2 Contraintes du projet

Afin de répondre de manière pertinente aux objectifs définis, le développement du système de recommandation devra prendre en compte un certain nombre de contraintes, à la fois techniques, métier et liées aux parties prenantes. Les principales contraintes identifiées sont :

Contraintes techniques :

- Traitement en temps réel des données utilisateur / base de données des produits
- Nécessité de traitements légers pour une expérience utilisateur fluide sur mobile comme sur ordinateur (temps de réponse optimale).
- Se connecter avec le système Shopify existant pour la récupération des données de stock, vente et catalogue produit.

Contraintes métier :

- Renouvellement fréquent de l'offre produits avec l'ajout de nouveaux designs et collections toutes les semaines.
- Stock parfois limité ou écoulable pour certains produits à durée de vie courte (ex : saisons, événements).

Enjeux business :

- Fidéliser la clientèle sur un marché attractif mais très concurrentiel.
- Booster la notoriété de la marque grâce à des ventes accrues et du référencement naturel.

- Maintenir la rentabilité de l'entreprise en augmentant le panier moyen et le ROI.

Enjeux utilisateur :

- Proposer une expérience d'achat sur-mesure et ludique à chaque visite.
- Répondre de façon proactive à leurs besoins par des suggestions personnalisées.

2.5 Backlog produit

Une fois que le contexte et les objectifs du projet ont été définis, il est essentiel d'établir le plan de développement de manière structurée. Le "product backlog", élément clé de la méthodologie agile SCRUM, permet justement de répertorier et ordonner l'ensemble des tâches à réaliser. Le tableau ci-dessous présente le Backlog de produit relatif à notre projet.

2.5 Backlog produit

ID	Feature	User Story	Priorité	Durée estimée (en semaines)
1	Collecte des données des produits	En tant qu'administrateur du système, je souhaite pouvoir collecter automatiquement les données des produits à partir de Shopify pour maintenir une base de produits à jour et fournir des recommandations précises.	Haute	2
2	Collecte des données d'utilisateurs	En tant qu'administrateur du système, je veux pouvoir configurer les champs de collecte de données utilisateur afin de recueillir les préférences des utilisateurs, ce qui permettra de personnaliser les recommandations basées sur leurs réponses aux questions.	Haute	1
3	Analyse des données collectées	En tant qu'administrateur du système, je veux pouvoir utiliser des techniques d'apprentissage automatique pour extraire des informations utiles à partir des données des utilisateurs, afin d'améliorer la précision des recommandations.	Haute	3
4	Développement de l'algorithme de recommandation	En tant que développeur, je veux concevoir et mettre en place un algorithme de recommandation qui utilise les données d'utilisateurs pour générer des recommandations personnalisées basées sur leurs préférences.	Haute	4
5	Conception de l'interface utilisateur	En tant que concepteur, je veux créer une interface utilisateur conviviale qui permet aux utilisateurs de visualiser et d'interagir facilement avec les recommandations personnalisées.	Haute	2
6	Test et amélioration	En tant que développeur, je veux prendre en compte les résultats des tests et les feedbacks des utilisateurs pour améliorer l'algorithme de recommandation, l'interface utilisateur et l'efficacité globale du système.	Moyenne	2

2.6 Planification des sprints

Le terme "SPRINT" est conceptualisé par la méthode "SCRUM". Un sprint indique le cycle de développement au cours duquel vont s'enchaîner un ensemble de tâches qui peuvent durer entre deux et quatre semaines. La figure 2.1 présente la planification des "Releases" et des "Sprints" de notre projet.

Sprint ID	Release 1: Développement de l'algorithme de recommandation	Release 2:Développement de l'Interface web
Sprint 0	Etude Technique	
Sprint 1	<div>1- Collecter les données des produits à partir du shopify</div> <div>2- Nettoyer les données</div> <div>3-Élaboration des spécifications de l'algorithme de recommandation.</div> <div>4- Mise en place de l'infrastructure de collecte des données utilisateur.</div>	
Sprint 2	<div>1- Implémentation de l'algorithme de recommandation initial basé sur les données collectées.</div> <div>2- Intégration des mécanismes d'apprentissage automatique pour améliorer la précision des recommandations.</div>	

2.7 Conclusion

	3- Tests unitaires de l'algorithme de recommandation.	
Sprint 3	<ul style="list-style-type: none">-Optimisation de l'algorithme de recommandation en utilisant des ensembles de données de test.-Amélioration de la performance et de l'efficacité de l'algorithme.	
Sprint 4		<ul style="list-style-type: none">- Conception de l'interface utilisateur pour l'affichage des recommandations personnalisées.-Mise en place de la structure de base de l'interface web.
Sprint 5		<ul style="list-style-type: none">-Implémentation des fonctionnalités d'affichage des recommandations sur l'interface web.-Personnalisation de l'interface pour permettre aux utilisateurs de modifier leurs préférences(Réponses).

2.7 Conclusion

Ce deuxième chapitre nous a permis de poser le contexte du projet et de définir de manière globale les objectifs et contraintes du futur système de recommandation. Le chapitre suivant portera désormais sur l'analyse détaillée des données disponibles.

Exploration des données

3.1 Introduction

Ce chapitre détaille l'exploration des principales sources de données disponibles en vue de développer le système de recommandation. Tout d'abord, nous décrivons le domaine d'activité de l'entreprise et ses produits afin de cadrer le périmètre fonctionnel du futur système. Puis, nous listons les besoins fonctionnels et non fonctionnels de notre système ainsi que l'identification des acteurs. Ensuite, nous identifions les principaux objectifs métier et contraintes auxquels devra répondre le système de recommandation pour assurer sa réussite. Enfin, nous clôturons ce chapitre par le backlog du produit la planification des sprints.

3.2 Présentation des sources de données

Cette section présente les trois sources des données disponibles pour notre système.

3.2.1 Base des produits 3D Crystal

Pour constituer la base de données produits, nous avons eu recours à une application dédiée à la plateforme Shopify.

Il s'agit d'Ablestar Bulk Product Editor, un outil capable d'interroger en masse le catalogue produits hébergé sur Shopify. Cet éditeur massif de produits nous a permis d'extraire l'ensemble des métadonnées structurées sous la forme d'un unique fichier au format CSV.



FIGURE 3.1 – Application Shopify : Ablestar Bulk Product Editor

Grâce à cette solution développée spécifiquement pour Shopify, nous avons pu récupérer de manière systématique et automatisée les détails de tous les produits référencés dans la boutique en ligne. Le fichier généré contient l'inventaire produits détaillé qui sera exploité pour le développement du système de recommandation.

3.2.2 Ensemble de questions-réponses préétablies

Afin de mieux comprendre les préférences des utilisateurs, une série de questions leur est posée lors de leur navigation sur le site. L'analyse de ces questionnaires personnalisés fournira des informations essentielles pour le système de recommandation. Le site 3D Crystal s'appuie sur un ensemble de questions types, élaborées en amont, afin de collecter les avis et préférences des visiteurs de manière structurée. Ces questions préétablies, au nombre de 5, portent sur des thématiques suivantes :

1. **Emotions**
2. **Occasion**
3. **Centre d'intérêts**
4. **Audience**
5. **Style personnel**

Le tableau ci-dessous présente de façon structurée les thématiques abordées dans le questionnaire, les questions correspondantes ainsi que les réponses possibles pour chaque question :

Thème	Question	Réponses
Émotions souhaitées	What emotions or feelings would you like shown in your gift ?	Love and romance;Happiness and joy;Peace and tranquility; Inspiration and motivation; Sentimental and nostalgic
Occasion du cadeau	What is the occasion/event you need a gift for?	Birthday;Anniversary; Housewarming;Holiday; celebration
Centres d'intérêt du receveur	Could you please tell me about the hobbies and interests of receivers?	Animals;Nature;Inspiring quotes; Art/Design;Constructions;Zodiac
Public cible	Is the gift targeted:	Child Audience;Teen Audience; Adult Audience;Senior Audience
Style personnel	How would you describe your personal style and aesthetic preferences?	"Deconstructionist Design" "Motorhead Chic" "Spiritual Stylista" "Narrative Garb" "Artisan Aesthetic" "Anatomical Appeal" "Symbolic Style" "Cosmic Chic" "Flora Fashion" "Animal Admirer"

3.2.3 Données d'entrée utilisateur pour le NLP

Deux sources principales de données textuelles sont collectées auprès des utilisateurs :

1. **Réponses aux questionnaires** Les réponses standardisées aux questions permettent de comprendre directement certaines préférences comme le style, les thématiques d'intérêt, les émotions souhaitées pour le cadeau, etc.
2. **Descriptions libres du profil et des besoins** Nous offrons un champ du texte libre au utilisateur pour décrire ses choix préféré afin d'extraire ces informations :
 - **Description du profil du receveur** : L'utilisateur est invité à renseigner librement les centres d'intérêt, la personnalité et le style esthétique préféré de la personne pour laquelle le cadeau est destiné.
 - **Expression des besoins et attentes pour le cadeau** : Un champ permet à l'utilisateur de décrire plus précisément le type de cadeau recherché, les éléments importants à prendre en compte, etc.

- **Réponses textuelles à certaines questions** : Pour certaines questions, des réponses sous forme de texte libre peuvent être fournies.

3.3 Analyse des données

3.3.1 Analyse du fichier CSV des produits 3D Crystal

Le fichier CSV regroupant l'inventaire produits est une source clé pour le développement du système de recommandation. Son exploration approfondie est nécessaire pour paramétrer l'algorithme de recommandation.

3.3.1.1 Structure des données

Le fichier CSV contient l'inventaire de l'ensemble des produits référencés sur le site de 3D Crystal.

Les principales colonnes identifiées sont :

1. **Product_id** : identifiant unique de chaque produit.
2. **Title** : titre ou nom du produit.
3. **Description** : description textuelle détaillée du produit.
4. **Tags** : étiquettes/mots-clés associés au produit (catégories, matériaux, motifs...)
5. **Images** : liens vers les images du produit en haute résolution.
6. **Website_url** : lien web vers la page détaillée du produit sur le site.
7. **Shopify_url** : lien e-commerce vers la fiche produit dans la boutique Shopify.
8. **Metafields** : autres métadonnées structurées sur le produit (prix, stock, options...)

Le fichier contient actuellement un millier de lignes, chaque ligne décrivant un produit de manière structurée via les attributs listés. L'exploration préliminaire a permis de constater que certaines colonnes comme les tags ou la description contiennent des données manquantes pour certains produits. Ce jeu de données fournit les informations clés sur chaque produit qui seront utilisées par le système de recommandation, notamment pour le calcul de similarité entre produits.

3.3.1.2 Évaluation de la qualité des données

L'évaluation de la qualité des données est essentielle pour garantir la fiabilité et la cohérence des informations contenues dans le jeu de données. Dans cette section, nous présentons une analyse approfondie de la qualité des données, en examinant plusieurs aspects clés.

Taux de données manquantes par variable :

- Certains attributs comme les tags sont manquants pour $<10\%$ des produits
- La description est complète à $>90\%$

Détection de valeurs aberrantes :

- Certains Produits avec des prix négatifs ou nuls (Produits de test)

Cohérence entre attributs :

- Correspondance entre le nombre d'images et de liens : On a besoin d'une seule image et d'un seul lien, donc on conserve l'image principale et on élimine les autres images supplémentaires.
- Adéquation des tags avec la description : Chaque produit a au moins un seul tag unique par rapport aux autres produits

Valeurs anormales

- Descriptions trop longues ou non informatives : Toutes les descriptions sont générées par les modèles d'intelligence artificielle GPT pour s'assurer que toutes les descriptions ont à peu près la même longueur et le même niveau de détail que le titre des produits.

Risques identifiés

- Recommandations non pertinentes avec les tags manquants
- Perte de sens avec les descriptions trop vagues
- Les produits non filtré par catégorie

3.3.2 Étude des questions-réponses

Après avoir analysé les données produites et compris la structure du catalogue, intéressons-nous à présent à la seconde source clé d'informations pour le système de recommandation : les réponses des utilisateurs aux questionnaires. Ces données subjectives fournies directement par les clients sont essentielles pour personnaliser l'expérience et affiner les recommandations. Leur étude approfondie est donc nécessaire.

3.3.2.1 Format des questions

L'ensemble des questions proposées dans le questionnaire utilisent un format à choix multiples, avec des réponses prédéfinies.

Cette approche présente plusieurs avantages :

- Elle simplifie et fluidifie l'expérience utilisateur en limitant le temps de réponse.
- Elle garantit l'homogénéité et la standardisation des réponses collectées.
- Elle facilite le traitement automatisé des données issues des questionnaires.
- Elle permet d'obtenir directement des profils utilisateurs formalisés sous forme codée.
- Elle limite les non-réponses grâce à l'exhaustivité des propositions de réponses.

Ce format de questions à choix multiples a donc été privilégié afin d'optimiser à la fois la qualité des données recueillies et l'ergonomie du questionnaire.

3.3.2.2 Objectifs et rôle dans le système de recommandation

Les réponses fournies ont plusieurs finalités :

- Mieux cerner les attributs déterminants pour la satisfaction de l'utilisateur.
- Identifier des segments de clientèle aux attentes similaires.
- Personnaliser le contenu et l'ordre de présentation des recommandations.
- Affiner en continu le modèle de recommandation grâce aux nouvelles réponses collectées.
- Améliorer l'expérience utilisateur en la rendant plus contextuelle.

L'analyse des questions-réponses est donc essentielle pour orienter le développement du système de manière pertinente.

3.3.3 Préparation des données textuelles pour la recommandation

Après La réception des entrées text par l'utilisateur il faut d'abord pré-traiter les données pour le nettoyage afin d'analyser les choix des utilisateurs (occasion,emotions,interets)

3.3.3.1 Objectifs du prétraitement

Le prétraitement vise à structurer ces données texte pour l'analyse par NLP afin d'en extraire :

- Les termes-clés décrivant les préférences implicites
- Les relations sémantiques entre produits et thématiques
- La polarité (positif/négatif) sur certains attributs

3.3.4 Étapes du prétraitement

Ces étapes visent à préparer les textes bruts pour l'analyse en appliquant notamment :

- Conversion en minuscules pour uniformiser et faciliter le traitement (exemple : "Cadeau" devient "cadeau")
- Suppression des accents et des caractères spéciaux qui peuvent poser problème (exemple : "café" devient "cafe")
- Elimination des retours à la ligne et des espaces multiples qui séparent artificiellement les termes
- Correction orthographique des mots mal écrits en se basant sur un dictionnaire (exemple : "interressé" devient "intéressé")
- Conservation des mots même incorrects si présents dans le vocabulaire produit
- Suppression des mots vides comme "le", "la", "des" n'apportant pas de sens

3.4 Conclusion

Cette phase essentielle de compréhension du domaine et des données à disposition jette les bases pour la modélisation et le développement du système de recommandation répondant de manière optimale aux besoins exprimés. Les étapes suivantes consisteront à formaliser les objectifs métiers puis à concevoir et valider l'architecture technique du système.

Modélisation et développement

4.1 Introduction

Après l'étape importante d'exploration et d'analyse des données présentée dans le chapitre précédent, nous allons à présent entamer la phase de modélisation et de développement du système de recommandation. Cette étape clé va nous permettre de formaliser les besoins métiers exprimés afin de concevoir une architecture technique répondant aux objectifs fixés. Nous nous appuyons pour cela sur le cadre méthodologique CRISP-DM qui structure le processus en différentes étapes.

Tout d'abord, nous modéliserons précisément les connaissances acquises lors de la phase d'étude des données. Puis nous spécifierons les exigences fonctionnelles et non-fonctionnelles du système à développer. Sur la base de ce cahier des charges, nous concevrons une première proposition d'architecture technique. Celle-ci sera ensuite développée de manière itérative selon la méthode agile. Enfin, nous mettrons en place une démarche de validation pour s'assurer que le système répond bien aux objectifs métiers fixés initialement.

Ce chapitre a pour objectif de poser les bases de la conception et du développement proprement dit du système de recommandation.

4.2 Modélisation du recommandation par questions-réponses

4.2.1 Méthodes utilisées

Lors du développement du notre système deux méthodes sont envisagées :

4.2.1.1 Similarité cosinus (Cousine Similarity)

Après transformation en sacs de mots avec TF-IDF, elle mesure la similarité entre deux vecteurs en calculant le cosinus de l'angle formé. C'est une mesure de similarité adaptée aux données textuelles. Vous trouvez dans la figure ci-dessous le process réalisé pour cette méthode :

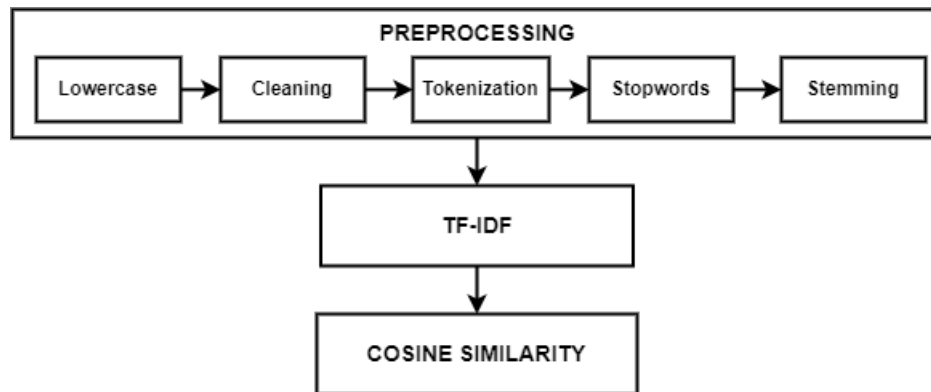


FIGURE 4.1 – Processus de la méthode Similarité cosinus

Fonctionnement de la similarité cosinus

La similarité cosinus mesure la similitude entre deux vecteurs représentant du texte. Elle se calcule comme le cosinus de l'angle formé par les deux vecteurs dans l'espace vectoriel. Plus cet angle est aigu, plus les vecteurs pointent dans la même direction et les textes sont similaires sémantiquement. La similarité cosinus renvoie donc une valeur comprise entre 0 (textes totalement dissimilaires) et 1 (textes identiques).

Avantages de la similarité cosinus

- Elle ne nécessite pas de normalisation des vecteurs, ce qui la rend robuste aux variations d'échelles.
- Le résultat est facilement interprétable comme un score de similitude entre 0 et 1.
- Elle est rapide à calculer même sur de grands corpus, ce qui permet son application à de larges jeux de données.
- Elle gère naturellement la polysémie grâce à la représentation vectorielle enrichie des contextes par le modèle TF-IDF.

Inconvénients de la similarité cosinus

- Elle ne prend pas en compte l'ordre des termes dans le texte, ce qui peut manquer des relations sémantiques.
- Les vecteurs creux générés posent des problèmes de stabilité numérique lors du calcul des angles sur de grands espaces.
- Le résultat dépend fortement de la qualité de la représentation vectorielle initiale obtenue avec TF-IDF.

- Elle ne convient pas aux données de type séquentiel comme les séries temporelles.

4.2.1.2 K plus proches voisins (KNN)

Elle consiste à recommander les k éléments les plus similaires dans la base en se basant sur une distance (euclidienne souvent).

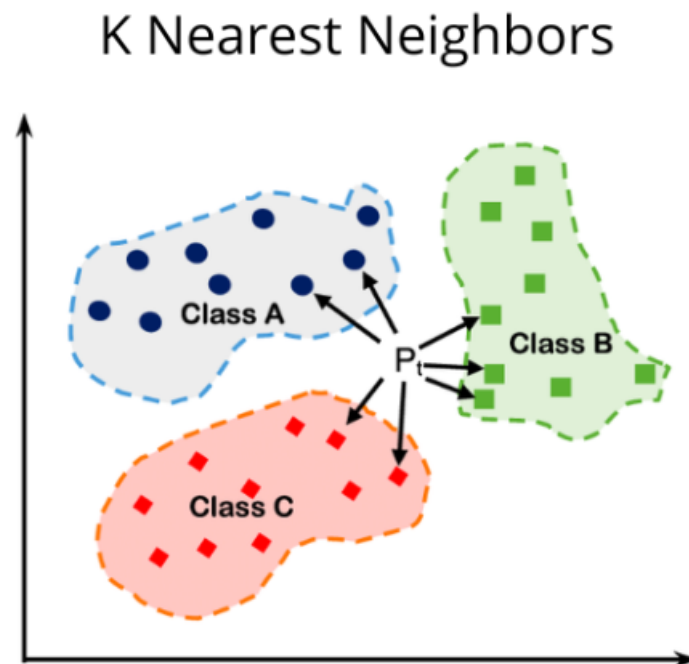


FIGURE 4.2 – Simulation de l’algorithme KNN

Fonctionnement de KNN

KNN classe un objet selon l’étiquette majoritaire parmi ses k plus proches voisins dans l’espace des caractéristiques, en se basant sur une mesure de distance (généralement euclidienne). Plus k est grand, plus la décision est robuste au bruit mais moins le modèle est réactif aux données locales.

Avantages de KNN

- Simple à mettre en œuvre et à interpréter, sans nécessiter d’apprentissage complexe.
- Efficace pour la classification et régression sur de nombreuses applications.

- Prend en compte la totalité des données d'apprentissage au moment de la prédiction.

Inconvénients de KNN

- Coûteux en temps de calcul et en espace mémoire lors de la recherche des k voisins, surtout pour de larges bases de données.
- Sensible au bruit et aux valeurs aberrantes dans les données d'apprentissage.
- Ne définit pas explicitement une frontière de décision claire comme d'autres algorithmes.
- Performances dégradées si la distribution des classes n'est pas uniforme dans l'espace des caractéristiques.
- Choix de la valeur optimale de k non trivial et dépendant des données.

4.2.2 Prétraitement des données

Avant toute mesure de similarité, les données textuelles doivent être pré-traitées :

Nettoyage du texte : suppression balises HTML, entités, caractères non alphabétiques.

Normalisation : La normalisation consiste à homogénéiser le texte afin de réduire sa variabilité syntaxique. Cela inclut généralement :

- Le passage à minuscules de l'ensemble du texte. Cela gomme les variations liées à la casse.
- La suppression de la ponctuation et des chiffres. Cela supprime certaines variations dans la structure du texte.

La normalisation vise à standardiser la forme des mots tout en conservant leur sens, afin de faciliter les comparaisons ultérieures.

Lemmatisation ou racinisation :

- **La lemmatisation** identifie le lemme, ou forme de base, de chaque mot. Par exemple "sauter", "sautant", "sauté" seront ramenés au lemme "sauter".
- **La racinisation** quant à elle ne cherche qu'à tronquer les flexions ou préfixes des mots. Exemple : "aimable" devient "aim".

Filtrage des mots vides : éliminer les articles, pronoms, etc... peu porteurs de sens

Ces étapes homogénéisent le texte et réduisent sa taille pour améliorer les performances.

4.2.3 Transformation en sacs de mots

Le modèle TF-IDF permet de transformer le texte prétraité en vecteurs de dimension égale au nombre de termes uniques dans le corpus. Chaque élément du vecteur correspond à la fréquence d'apparition d'un terme, pondérée par sa rareté dans le reste du corpus. Cette représentation vectorielle du texte permet d'appliquer des mesures de similarité cosinus ou distances euclidiennes. La figure ci-dessous présente schématiquement le principe du modèle TF-IDF :

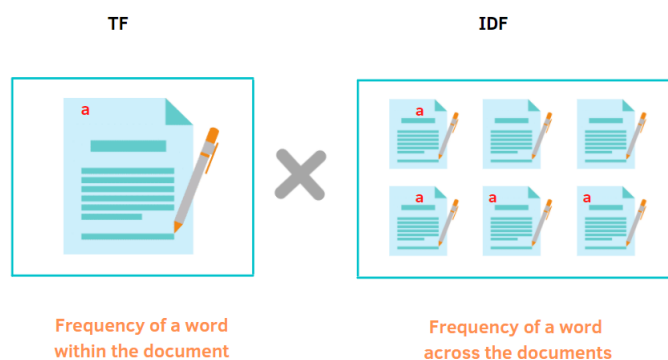


FIGURE 4.3 – principe du modèle TF-IDF

4.2.4 Application aux questions-réponses

Jusqu'à présent, nous avons présenté les méthodes théoriques utilisables pour le système de recommandation par questions-réponses, à savoir la similarité cosinus et KNN.

Dans notre cas d'usage, nous appliquons plus spécifiquement la similarité cosinus. L'utilisateur répond à un questionnaire sous forme de questions structurées, ce qui constitue son profil décrit de manière vectorielle. De même, les produits de la base de données sont décrits par des métadonnées textuelles vectorisées à l'aide du modèle TF-IDF précédemment introduit (cf section X).

Les étapes clés sont les suivantes :

1. Prétraitement du profil utilisateur et des descriptions produits pour homogénéiser les données textuelles.
2. Transformation en sacs de mots avec TF-IDF pour obtenir une représentation vectorielle des différents textes.
3. Calcul de la similarité cosinus, métrique adaptée à cette représentation vectorielle et présentée en section Y, entre le profil utilisateur et chaque produit.
4. Recommandation des Z produits présentant la plus forte similarité cosinus, soit les plus pertinents au regard du profil de l'utilisateur.

Cette approche exploite de manière unifiée les concepts théoriques précédemment introduits et les applique de façon concrète à notre cas d'usage particulier basé sur un questionnaire structuré.

4.3 Modélisation du recommandation par NLP

L'analyse des données textuelles collectées auprès des utilisateurs a mis en évidence la richesse d'informations qu'elles contiennent sur les préférences et le contexte d'achat. Il est désormais nécessaire d'extraire ces informations de manière structurée pour alimenter le système de recommandation.

Les techniques de traitement automatique du langage naturel (NLP) sont pertinentes pour modéliser les connaissances latentes contenues dans les textes. Cette section étudie les différentes approches NLP envisageables afin de sélectionner la mieux adaptée à nos données et objectifs.

4.3.1 Principe du NLP

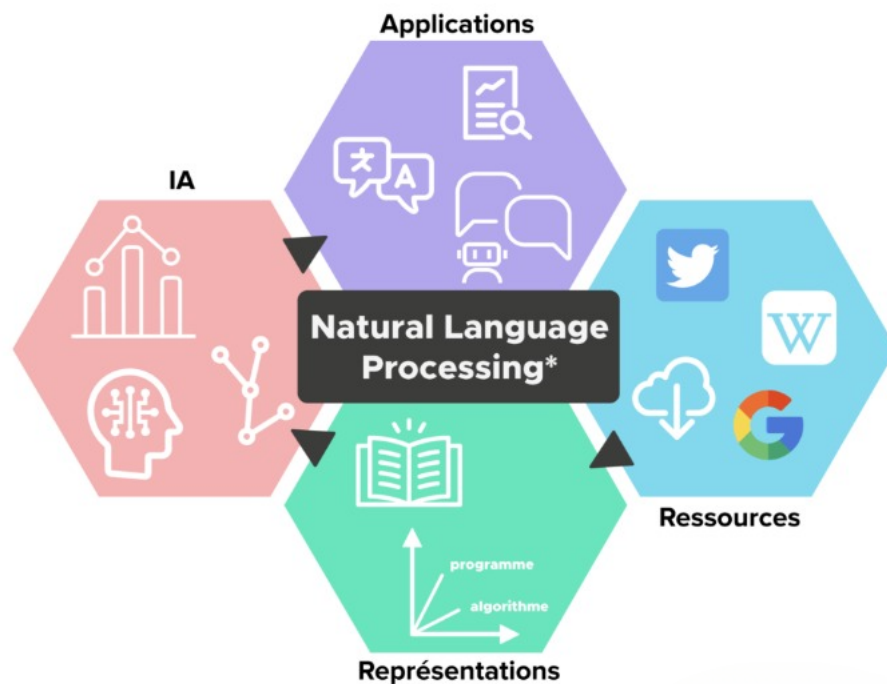
La discipline du NLP, également connue sous le nom de Traitement du Langage Naturel, se concentre principalement sur la compréhension, la manipulation et la création du langage naturel par les machines. De cette manière, le NLP représente véritablement un pont entre la science informatique et la linguistique. Il aborde donc la possibilité pour la machine d'interagir directement avec l'être humain.

Dans l'ensemble, on peut identifier deux éléments clés à tout problème de NLP :

La section « linguistique » implique de préparer et de convertir les informations en entrée en un jeu de données exploitable.

La section "apprentissage automatique" ou "Science des données" concerne l'utilisation de modèles d'apprentissage automatique ou de Deep Learning pour analyser ce jeu de données.

4.3.2 Principaux domaines d'application du NLP



**Traitement Automatique du Langage Naturel*

FIGURE 4.4 – Domaines d'application du NLP

Le NLP a permis des avancées significatives dans divers domaines. Il est utilisé pour traduire automatiquement des textes, analyser les sentiments des auteurs, cibler les consommateurs en marketing, développer des chatbots pour interagir avec les clients, classer et organiser des textes, extraire des informations de documents, corriger automatiquement l'orthographe et produire des résumés de texte, etc. Ces applications démontrent le potentiel du NLP pour automatiser et améliorer les tâches liées au langage naturel, offrant ainsi des opportunités dans de nombreux domaines professionnels.

Dans notre projet de recommandation de cadeaux basé sur le traitement du texte, nous exploitons ces avancées du NLP pour analyser les préférences des utilisateurs, comprendre leurs choix et leur fournir des recommandations de

cadeaux personnalisées. En utilisant des techniques de traitement du langage naturel, nous sommes en mesure d'extraire des informations pertinentes des descriptions et des préférences des utilisateurs, ce qui nous permet de proposer des cadeaux adaptés à leurs goûts et à leurs besoins. Le NLP joue un rôle crucial dans notre projet en permettant une compréhension fine du langage naturel et en facilitant la recommandation de cadeaux personnalisés pour une expérience utilisateur optimale.

4.3.3 Les principaux algorithmes NLP

Cette section présente un tableau comparatif décrit les principaux algorithmes NLP, en précisant leur fonctionnement, forces et limites. Cela nous permettra d'identifier les candidats les plus prometteurs :

4.3 Modélisation du recommandation par NLP

Algorithme	Méthode d'entraînement	Méthodologie d'évaluation	Forces	Faiblesses
BERT	Apprentissage transféré en utilisant le pré-entraînement du modèle de langage masqué	Évalué sur diverses tâches de référence GLUE incluant question-réponse, analyse de sentiment etc.	Meilleurs résultats de l'état de l'art. Tire parti des grandes données non étiquetées pour le pré-entraînement. Peut être facilement ré-entraîné pour de nouvelles tâches.	Nécessite d'importantes ressources de calcul pour le pré-entraînement. Difficile à entraîner à partir de zéro.
Classifieur zéro-shot	Apprentissage transféré en utilisant les plongements BERT/XLNet	Évalué sur la capacité à effectuer la classification d'intentions et la génération de réponses directement à partir du texte sans ré-entraînement	Ne nécessite aucune donnée étiquetée pour la tâche cible. Peut généraliser à des classes non vues	Limité aux tâches de classification où les relations entre les entrées et les sorties sont très explicites

RNN	Entraîné à partir de zéro sur des données étiquetées	Évalué sur la tâche NER en utilisant les métriques standards comme la précision, le rappel, le F1-score	Efficace pour la modélisation de séquences comme NER grâce à l'utilisation du contexte	Architecture complexe. Nécessite de grandes données étiquetées et de calculs pour l'entraînement
DNN	Entraîné à partir de zéro sur des données étiquetées	Évalué sur les tâches de classification de texte en utilisant la précision et les métriques de perte	Architecture simple. Facile à entraîner et à interpréter	Ne capture pas efficacement l'information de séquence/contexte. Nécessite de grandes données étiquetées

4.3.4 Classifieur zéro-shot

4.3.4.1 Fonctionnement du Classifieur zéro-shot

Les classifieurs zéro-shot permettent de classer des données dans des catégories non vues à l'entraînement, sans exemples étiquetés pour ces nouvelles classes.

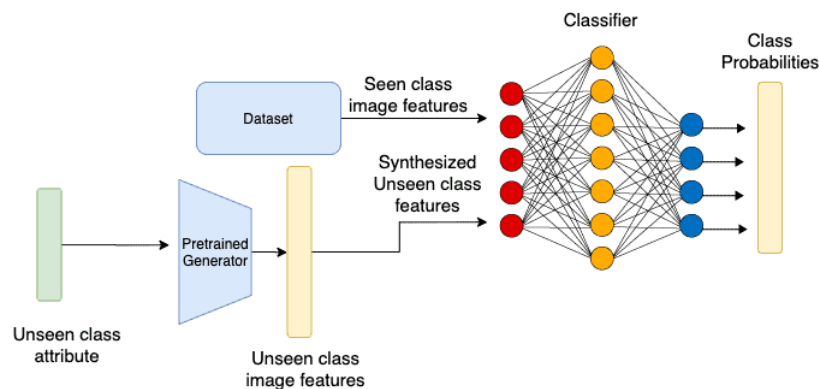


FIGURE 4.5 – Zero-Shot Learning

Notre système met en œuvre cette approche en deux étapes clés :

1. Apprentissage des plongements lexicaux

Nous entraînons un modèle de plongement de mots non supervisé (Word2Vec par exemple) sur un grand corpus généraliste. Cet apprentissage projette chaque mot dans un espace vectoriel où les similarités sémantiques sont préservées. Il capture les régularités distributionnelles sans supervision.

2. Classification zéro-shot Pour chaque nouveau texte d'entrée :

- Nous encodons son contenu et les labels candidats dans l'espace des plongements.
- Un calcul de similarité cosinus est effectué entre chaque représentation de label et le texte.
- Le label dont le vecteur de plongement est le plus proche du texte (cosinus maximal) est assigné.

Cette approche très peu supervisée permet de catégoriser les réponses de l'utilisateur dans les classes existantes de façon non supervisée. Les prédictions sont ensuite transformées en sacs de mots avec TF-IDF, à l'instar des descriptions de produits. La similarité cosinus entre ces représentations vectorielles permet enfin de recommander les items les plus pertinents de manière contextualisée.

Notre système exploite judicieusement les capacités zéro-shot et la représentation vectorielle du langage pour formuler des recommandations adaptées.

4.4 Conclusion

Ce chapitre a jeté les bases de la conception et du développement du système de recommandation en alignant les besoins métiers avec une architecture technique appropriée. En suivant une approche méthodique et itérative, nous avons pu créer un système de recommandation robuste et fiable, capable de répondre aux attentes des utilisateurs et de contribuer aux objectifs de l'entreprise.

Evaluation et Déploiement

5.1 Introduction

Ce chapitre se concentre sur l'évaluation et le déploiement des résultats obtenus dans notre projet de recommandation, en mettant en avant l'importance de mesurer les performances du modèle développé et de l'intégrer dans un environnement réel pour une utilisation pratique, tout en considérant les aspects techniques, la formation des utilisateurs et l'évaluation continue des résultats déployés.

5.2 Critères d'évaluation théorique

Au cours de notre processus d'évaluation nous mettons en compte ces critères :

Performance et temps de réponse

- Assurer que le système de recommandation répond rapidement aux requêtes des utilisateurs, en optimisant les algorithmes et les opérations de traitement des données.
- Utiliser des techniques de mise en cache pour accélérer les réponses et minimiser les temps de chargement.
- Effectuer des tests de charge et d'évolutivité pour garantir que le système peut gérer une charge importante sans compromettre ses performances.

Convivialité et expérience utilisateur

- Concevoir une interface utilisateur intuitive et conviviale pour faciliter la navigation et l'interaction avec le système de recommandation.
- Personnaliser les recommandations en fonction des préférences et des comportements des utilisateurs pour améliorer leur expérience.
- Fournir des options de filtrage et de tri pour permettre aux utilisateurs de trouver facilement les recommandations qui les intéressent.

Sécurité des données utilisateur

- Mettre en place des mesures de sécurité appropriées pour protéger les données des utilisateurs, notamment en utilisant des protocoles de chiffrement et en suivant les meilleures pratiques en matière de sécurité.
- Obtenir le consentement des utilisateurs et respectez les réglementations en vigueur concernant la confidentialité des données personnelles.

Évolutivité et extensibilité du système

- Concevoir le système de recommandation de manière modulaire et extensible, en utilisant des architectures flexibles et évolutives.
- Prévoir des mécanismes pour ajouter de nouveaux types de recommandations ou étendre les fonctionnalités existantes sans compromettre la stabilité du système.

Performance Technique

- Infrastructure adaptée : Utiliser des ressources matérielles (RAM et CPU Performants) et logicielles (Visual Studio - Google Collab) adaptées pour exécuter nos algorithmes efficacement, en tenant compte de la charge attendue et des besoins en calcul intensif.
- Temp d'apprentissage : Utilisez des modèles pré-entraînés puissants tels que BERT ([facebook/bart-large](https://huggingface.co/facebook/bart-large)) pour bénéficier d'une meilleure performance de classification

5.3 Cycle de vie du projet



FIGURE 5.1 – Cycle de vie du projet

L'évaluation joue un rôle essentiel dans le cycle de vie d'un projet de data science, surtout pour "Comprendre les besoins clients". Elle mesure l'efficacité, la performance et la validité des modèles développés, et assure leur conformité aux objectifs du projet et aux exigences des parties prenantes.

Elle quantifie objectivement la capacité du modèle à résoudre les problèmes identifiés et à fournir des résultats fiables. Cela nécessite la sélection de mesures adéquates, telles que l'exactitude, la précision, le rappel, la F-mesure, etc., en fonction de la nature du problème et des objectifs spécifiques. Pour assurer une évaluation efficace, il est crucial d'employer des méthodologies rigoureuses afin de garantir des résultats fiables et généralisables.

Dans la section suivante, nous détaillerons la méthodologie employée pour mener à bien cette évaluation.

5.4 Méthodologie d'évaluation

L'évaluation de la classification de texte sans apprentissage est un aspect essentiel pour mesurer la qualité et la performance des modèles de classification. Dans le contexte de la Zero-Shot Text Classification, l'objectif est de prédire les classes de texte sans avoir besoin d'exemples d'entraînement spécifiques à chaque classe. Cette approche offre une flexibilité et une généralisation accrues, mais nécessite une évaluation rigoureuse pour évaluer son efficacité.

5.4.1 Prédiction au niveau des phrases

L'évaluation de la classification de texte sans apprentissage repose sur la prédiction des classes au niveau des phrases. Les phrases sont traitées individuellement et les modèles attribuent des scores de confiance à chaque classe prédite. Cette approche permet une classification fine-grained et une analyse précise des prédictions.

5.4.2 Agrégation des prédictions au niveau des textes

Pour obtenir une prédiction globale pour un texte donné, il est nécessaire d'agréger les scores de confiance des phrases. Une approche courante consiste à calculer la moyenne des scores pour chaque classe prédite. Cette agrégation permet d'obtenir un score de confiance global pour chaque classe, qui peut être utilisé pour prendre la décision finale de classification.

5.4.3 Calcul de l'exactitude

Pour calculer l'exactitude, on utilise des bibliothèques telles que scikit-learn, qui offrent des fonctions prêtes à l'emploi pour évaluer les performances des modèles de classification. En utilisant les classes prédites et les classes réelles des échantillons d'évaluation, on peut facilement calculer l'exactitude à l'aide de la formule : $\text{exactitude} = (\text{nombre de prédictions correctes}) / (\text{nombre total d'échantillons})$.

5.5 Déploiement

La dernière étape du méthodologie CRISP-DM est le déploiement, où nous utiliserons diverses technologies pour créer une application web conviviale et facile à utiliser. Cette section examine les technologies utilisées pour le déploiement et les interfaces de notre application. Cela nous permettra de proposer une solution intégrée et efficace pour la reconnaissance automatique des plaques minéralogiques.

5.5.1 Technologies utilisées

. Python



FIGURE 5.2 – Logo Python

Python est un langage de programmation interprété, multiparadigme et multiplateformes. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet.

. Flask



FIGURE 5.3 – Logo Flask

Flask est un micro framework open-source de développement web en Python. Il est classé comme microframework car il est très léger. Flask a pour objectif de garder un noyau simple mais extensible.

. HTML



FIGURE 5.4 – Logo HTML

Le HyperText Markup Language, généralement abrégé HTML ou, dans sa dernière version, HTML5, est le langage de balisage conçu pour représenter les pages web.

. CSS



FIGURE 5.5 – Logo CSS

Les feuilles de style en cascade, généralement appelées CSS de l'anglais Cascading Style Sheets, forment un langage informatique qui décrit la présentation des documents HTML et XML. Les standards définissant CSS sont publiés par le World Wide Web Consortium.

. JavaScript



FIGURE 5.6 – Logo Javascript

JavaScript est un langage de programmation de scripts principalement employé dans les pages web interactives.

Avec les langages HTML et CSS, JavaScript sont au cœur des langages utilisées par les développeurs web.

5.5.2 Interfaces

Cette partie représente la phase de mise en œuvre de l'application. Nous exposons alors les différentes interfaces de notre solution proposée.

. Page d'accueil

Lorsque vous utilisez notre application Web, vous serez redirigé vers notre page d'accueil qui propose deux interfaces utilisateur. Ces interfaces vous permettent d'interagir avec notre système de recommandation de cadeaux en cristal 3D en utilisant soit un questionnaire interactif, soit une description du destinataire du cadeau.

La page d'accueil est représentée dans l'illustration ci-dessous :

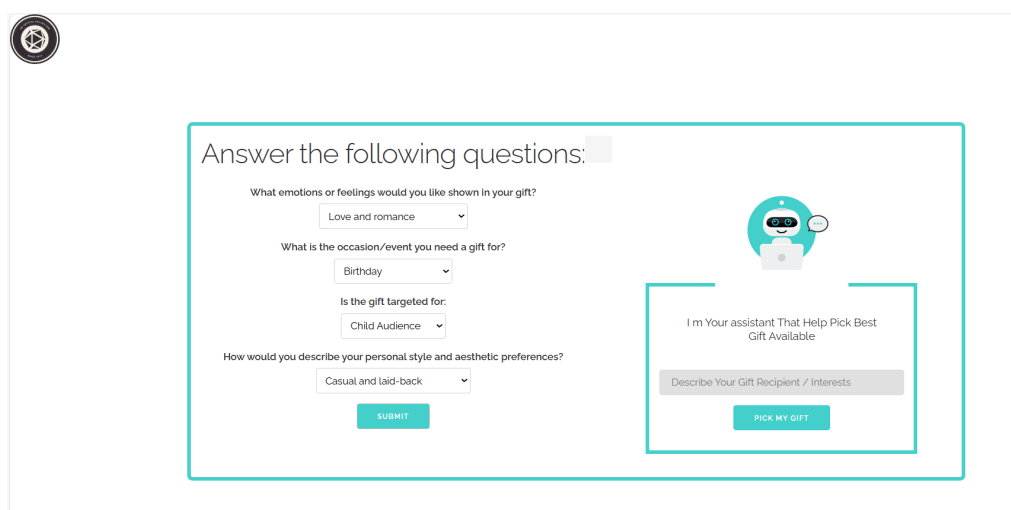
The screenshot shows a web interface for a gift recommendation system. On the left, there is a questionnaire titled 'Answer the following questions:'. It contains four questions with dropdown menus: 'What emotions or feelings would you like shown in your gift?' (selected 'Love and romance'), 'What is the occasion/event you need a gift for?' (selected 'Birthday'), 'Is the gift targeted for:' (selected 'Child Audience'), and 'How would you describe your personal style and aesthetic preferences?' (selected 'Casual and laid-back'). A 'SUBMIT' button is at the bottom of the questionnaire. On the right, there is a section titled 'I'm Your assistant That Help Pick Best Gift Available' featuring a cartoon robot icon. Below this, there is a text input field labeled 'Describe Your Gift Recipient / Interests' and a 'PICK MY GIFT' button.

FIGURE 5.7 – Page d'accueil

. Page de recommandation basée sur le traitement des questions/réponses

Une fois l'utilisateur fourni ses réponses aux questions, notre système entre en action. Il compare ses informations avec les titres et les descriptions des produits que nous avons préalablement stockés dans notre base de données. Grâce à des algorithmes de similarité, nous identifions les trois produits les plus pertinents et les plus similaires à ses préférences qui seront affichés dans la page.

La page de recommandation est représentée dans l'illustration ci-dessous :

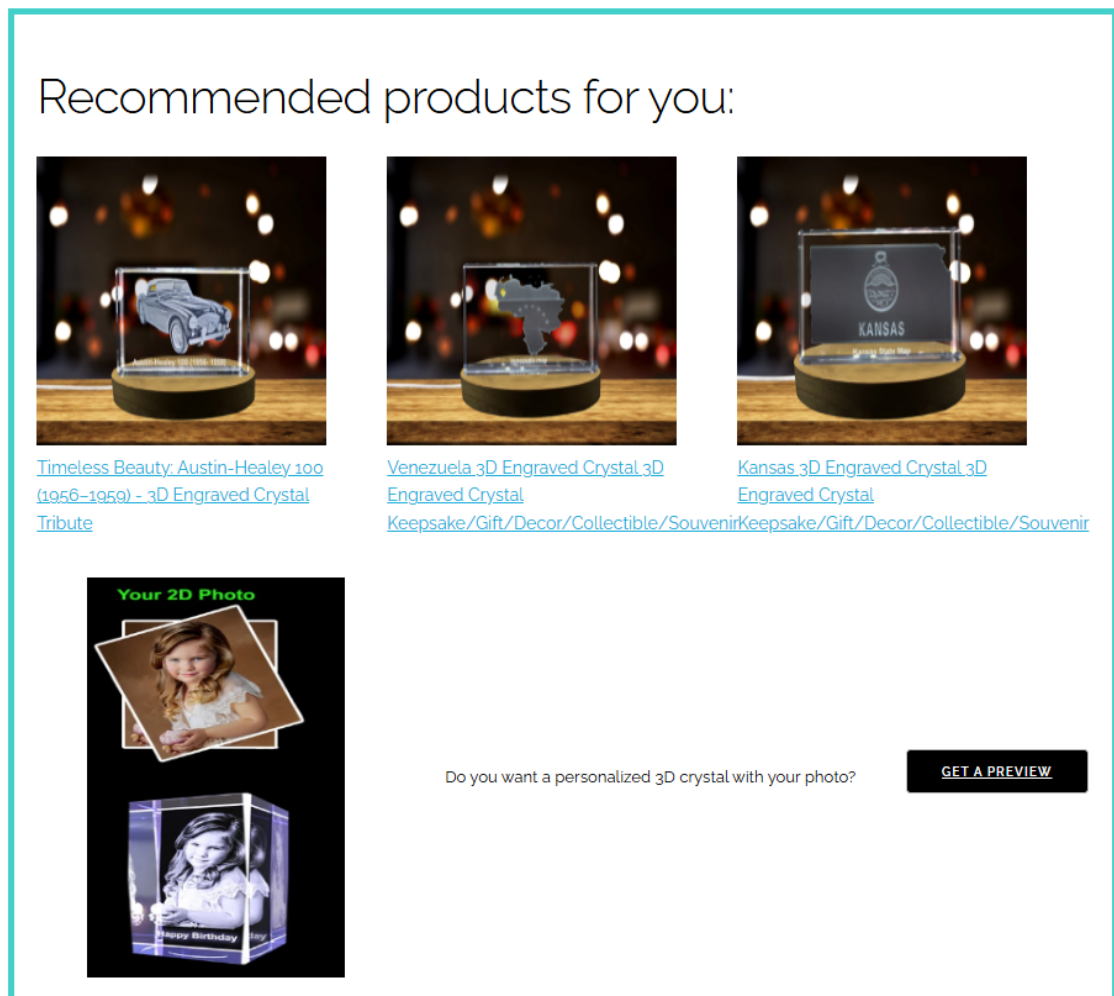


FIGURE 5.8 – Page de recommandation basée sur le traitement des questions/réponses

. Page de recommandation basée sur le traitement du text NLP

Notre page de recommandation basée sur le traitement du texte NLP vous offre une solution avancée pour trouver les produits qui correspondent à vos besoins. Après avoir soumis votre texte, notre système utilise des modèles de classification pré-entraînés pour prédire les catégories les plus appropriées à votre demande. Les classes prédites sont ensuite affichées, ce qui vous permet de comprendre comment votre texte a été interprété et catégorisé. Enfin, la page vous présente les trois meilleurs produits recommandés, soigneusement sélectionnés en fonction de leur pertinence par rapport à votre demande initiale.

La page de recommandation est représentée dans l'illustration ci-dessous :

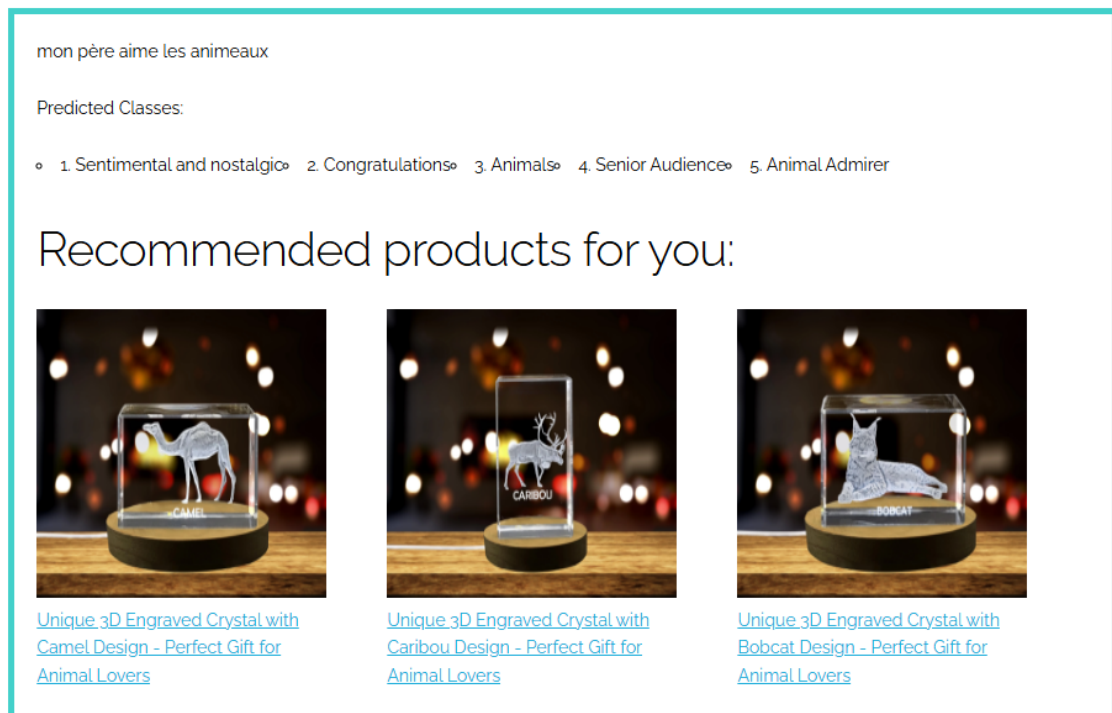


FIGURE 5.9 – Page de recommandation basée sur le traitement du text NLP

5.5.3 Hébergement

Après avoir terminé le déploiement de notre solution, nous avons choisi de l'héberger sur la plateforme Heroku.



FIGURE 5.10 – Logo Heroku

Heroku est un service d'hébergement dans le cloud qui offre une grande facilité de déploiement et de gestion des applications web. Il prend en charge les applications Flask, ce qui en fait un choix idéal pour notre application Flask.

L'hébergement sur Heroku présente plusieurs avantages :

- Heroku nous permet de mettre notre application en ligne rapidement et facilement. Heroku gère automatiquement les tâches liées aux serveurs, à la mise à l'échelle et à la gestion des ressources, ce qui nous permet de nous concentrer sur le développement de notre application.
- Heroku offre des fonctionnalités telles que le déploiement continu, ce qui signifie que nous pouvons mettre à jour notre application en toute transparence sans interruption de service.
- Accéder aux journaux d'application pour le débogage et la surveillance.

En hébergeant notre application Flask sur Heroku, nous bénéficions d'une infrastructure fiable, sécurisée et évolutive. Cela nous permet de fournir une expérience utilisateur fluide et d'assurer la disponibilité de notre application en tout temps.

5.6 Conclusion

En conclusion de ce chapitre consacré à l'évaluation et au déploiement de notre solution, nous avons réalisé une évaluation approfondie de notre système, en mettant l'accent sur sa performance, sa précision et sa convivialité. Grâce à ces tests rigoureux, nous avons pu améliorer et optimiser notre application pour offrir une expérience utilisateur optimale. De plus, nous avons choisi de déployer notre application en utilisant les technologies principales, notamment Flask comme framework de développement et Heroku comme plateforme d'hébergement.

Conclusion générale

Dans cette étude, nous avons présenté un projet visant à développer un système de recommandation pour les produits 3D Crystal. Nous avons commencé par introduire le cadre et le contexte du projet, en mettant en évidence la problématique existante et en décrivant la solution proposée ainsi que les objectifs à atteindre.

Ensuite, nous avons examiné en détail le domaine des produits 3D Crystal et analysé les besoins des utilisateurs. Nous avons identifié les acteurs impliqués, les besoins fonctionnels et non fonctionnels, ainsi que les objectifs et les contraintes du projet. Nous avons également établi un backlog produit et planifié les sprints pour guider le développement.

L'étape suivante a été consacrée à l'exploration des données disponibles. Nous avons examiné les sources de données, notamment la base de produits 3D Crystal, les questions-réponses préétablies et les données d'entrée utilisateur pour le traitement du langage naturel (NLP). Nous avons effectué une analyse approfondie de ces données, y compris leur prétraitement, afin de les préparer pour la recommandation.

Dans la section de modélisation et de développement, nous avons présenté deux approches de recommandation : une basée sur les questions-réponses et une basée sur le NLP. Nous avons décrit les méthodes utilisées, y compris le prétraitement des données et la transformation en sacs de mots. De plus, nous avons exploré les principes du NLP et les algorithmes associés, en mettant en évidence l'utilisation d'un classifieur zéro-shot.

La phase d'évaluation et de déploiement a été abordée dans la section suivante. Nous avons défini les critères d'évaluation théorique, établi le cycle de vie du projet et présenté une méthodologie d'évaluation. Nous avons également discuté du déploiement du système, en détaillant les technologies utilisées, les interfaces mises en place et les aspects liés à l'hébergement.

Bien que nous ayons développé un système de recommandation fonctionnel pour les produits 3D Crystal, il existe encore des possibilités d'expansion

et d'amélioration. Par exemple, nous pourrions envisager d'intégrer des fonctionnalités supplémentaires telles que des recommandations basées sur les évaluations des utilisateurs, des recommandations contextuelles en fonction des tendances actuelles ou des recommandations croisées avec d'autres produits connexes.

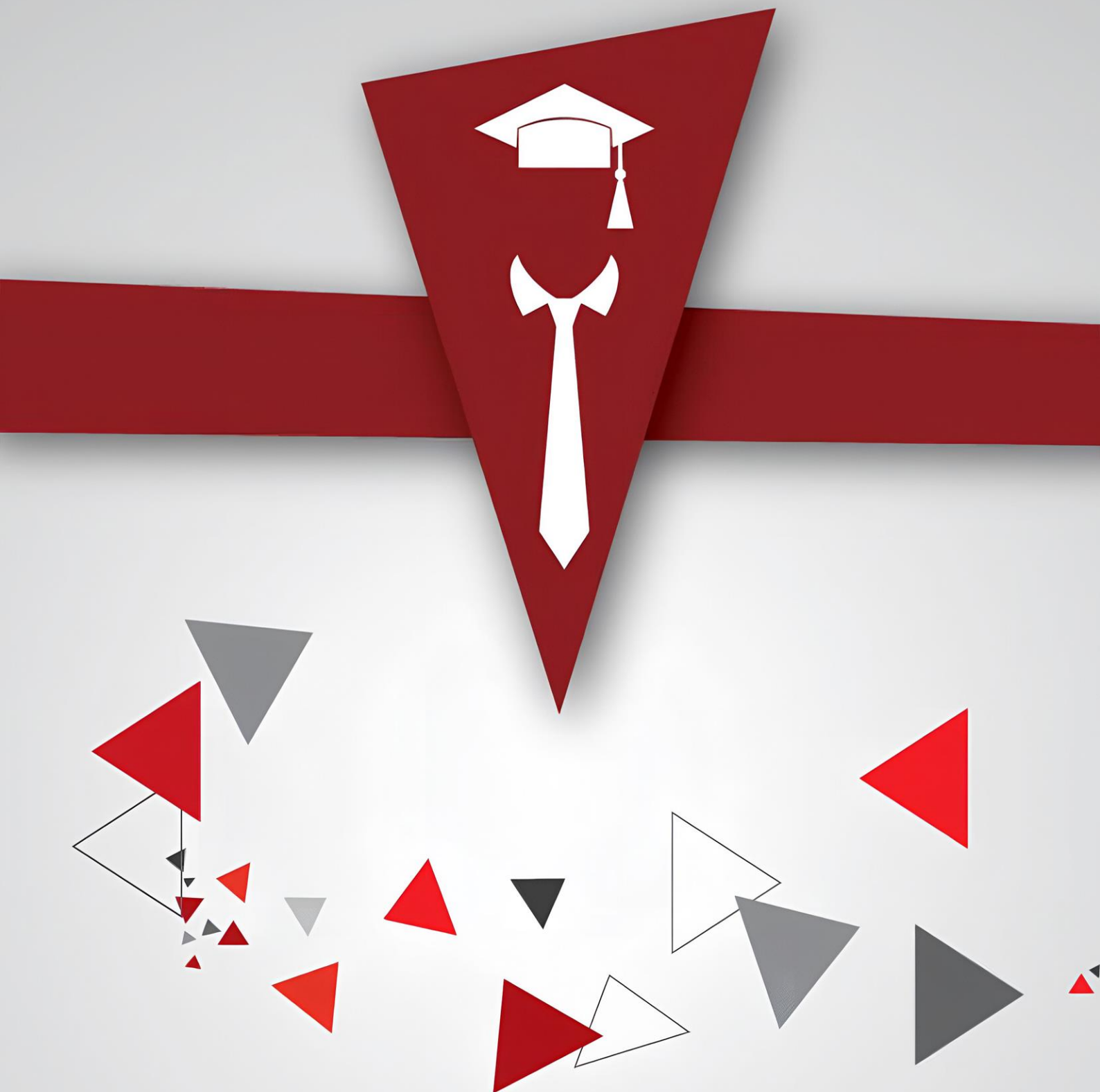
En conclusion, ce projet de développement d'un système de recommandation pour les produits 3D Crystal représente une contribution significative dans le domaine de l'intelligence artificielle et du traitement du langage naturel. Les différentes étapes, de l'analyse des besoins à la modélisation et au déploiement, ont été abordées de manière approfondie, offrant ainsi une base solide pour la mise en œuvre du système. Les résultats obtenus à partir de l'évaluation théorique et les perspectives d'amélioration future montrent le potentiel et l'importance de ce projet dans le domaine des recommandations personnalisées.

bibliographie

ANNEXE A

ANNEXE B

ANNEXE C



ECOLE SUPÉRIEURE PRIVÉE D'INGÉNIEURS DE MONASTIR

www.esprim.tn - E-mail : contact@esprim.tn

Adresse : Boulevard de l'environnement BP 104, Monastir République – 5060 Tunisie

Tél. : +216 94 840 200 - **Fax. :** +216 73 500 200
