



**ECOLE MAROCAINE DES
SCIENCES DE L'INGENIEUR**

Membre de
HONORIS UNITED UNIVERSITIES



Rapport du projet de fin d'année

3ème année Informatique et Réseaux

Sous le thème

**PROCEDURE D'ANALYSE DES SENTIMENTS
EN UTILISANT L'INTELLIGENCE ARTIFICIELLE**

Réalisé par :

Ayoub Majjid

Rochdi mohamed Amine

Encadré par :

Tuteur de l'école : Mme. BENSALAH

ANNEE UNIVERSITAIRE: 2023-2024

Remerciements

Nous tenons à exprimer notre profonde gratitude à toutes les personnes qui ont contribué à la réalisation de ce projet.

Tout d'abord, nous remercions sincèrement nos parents pour leur soutien inconditionnel. Leur encouragement constant et leur confiance en nous ont été des moteurs essentiels tout au long de cette aventure. Merci pour votre patience et votre compréhension, et pour avoir toujours cru en nos capacités.

Nous souhaitons également remercier nos camarades pour leurs propositions et idées additionnelles. Leur contribution a été précieuse et a permis d'enrichir notre projet. Merci pour votre aide, vos suggestions constructives et votre esprit de coopération. Vos contributions nous ont été d'une grande aide.

Enfin, nous tenons à exprimer notre reconnaissance à notre encadrant. Votre expertise, vos conseils avisés et votre disponibilité ont été des atouts inestimables. Merci pour votre soutien tout au long de ce projet, pour votre encadrement rigoureux et pour avoir toujours été à l'écoute. Votre guidance a été essentielle pour mener ce projet à bien.

À tous, merci infiniment.

Dédicace

Je dédie ce travail à ma famille, pour leur soutien inconditionnel, leur amour et leur encouragement constants. Leur présence et leur soutien ont été ma force tout au long de ce parcours académique. À mes parents, qui ont toujours cru en moi et m'ont encouragé à poursuivre mes rêves, je vous suis infiniment reconnaissant.

À mes amis et collègues, qui ont partagé les hauts et les bas de cette aventure, je vous remercie pour votre amitié, votre soutien et votre compréhension. Vos encouragements et votre présence ont rendu ce voyage plus joyeux et significatif.

Je souhaite également dédier ce travail à mes professeurs et mentors, dont la sagesse, les conseils précieux et l'expertise m'ont guidé à travers les défis académiques. Votre enseignement a été une source d'inspiration et a contribué à ma croissance personnelle et professionnelle.

Enfin, à toutes les personnes qui, d'une manière ou d'une autre, ont contribué à cette réussite, je vous exprime ma profonde gratitude. Ce travail est le fruit de nombreuses contributions et de soutiens précieux, et il est avec une immense reconnaissance que je le dédie à chacun d'entre vous.

Table des matières

Remerciements	3
Dédicace	4
Liste des figures	7
Liste des tableaux	8
Abstract	9
1 Introduction générale	10
1.1 Contexte	10
1.2 Organisation du mémoire	10
2 Analyse et Conception	12
Introduction	12
2.0.1 Description des besoins fonctionnels	12
2.0.2 Description des besoins ergonomiques	13
2.0.3 Description des besoins graphiques	13
2.0.4 Description des besoins techniques	13
2.0.5 Modélisation UML	13
Conclusion	13
3 Analyse des sentiments et revue de la littérature	24
Introduction	24
3.0.1 Sentiments et opinions	24
3.0.2 Traitement du langage naturel (NLP)	24
3.0.3 Analyse des sentiments	24
3.0.4 Revue de la littérature	25
Conclusion	25
4 Procédure d'analyse des sentiments	26
Introduction	26
4.0.1 Collection de données	26
4.0.2 Prétraitement de la base de données	27
4.0.3 Word Embedding (Vectorisation du texte)	27
4.0.4 Classification	30
Conclusion	36

5 Interfaces de l'application	38
Introduction	38
5.0.1 Les outils utilisés	38
5.0.2 Prétraitement de données (captures d'écran)	39
5.0.3 Word Embedding (captures d'écran)	39
5.0.4 Classification (captures d'écran)	40
5.0.5 Interfaces (captures d'écran)	40
Conclusion	41
Webographie	53

Table des figures

2.1	Diagramme D'utilisation de cas	14
2.2	Diagramme de séquence du consultation	15
2.3	Diagramme de séquence de home	16
2.4	Diagramme de séquence de cities	17
2.5	Diagramme de séquence d'about	18
2.6	Diagramme de séquence du login	19
2.7	Diagramme de séquence d'authentification	20
2.8	Diagramme de séquence du register	21
2.9	Diagramme de séquence du chatbot	22
2.10	Diagramme de séquence du profile	23
4.1	$Wt,Doc = TFt,Doc * \log(N/Nt)$	27
4.2	FastText	29
4.3	Techniques d'analyse des sentiments	31
4.4	Récapitulatif des méthodes classiques d'analyse des sentiments	32
4.5	Structure générale d'un réseau de neurones	33
4.6	Fonctionnement d'un réseau de neurones	34
4.7	Architecture d'un réseau de neurones convolutif	34
4.8	Convolution d'une matrice via un filtre	35
4.9	Architecture d'un seul neurone LSTM	36
5.1	Exemple de prétraitement des données	42
5.2	Exemple de classification avec régression logistique	43
5.3	Rapport Résultat du model LR	43
5.4	Exemple de classification avec Naive Bayes	44
5.5	Rapport Résultat du model Naive Bayes	45
5.6	Exemple de classification avec arbres de décision	46
5.7	Rapport Résultat du model arbres de décision DT	47
5.8	Exemple de classification avec SVM	48
5.9	Rapport Résultat du model SVM	49
5.10	Capture d'écran de la page d'accueil	49
5.11	Capture d'écran de la page cities	50
5.12	Capture d'écran de la page about	50
5.13	Capture d'écran de la page login	51
5.14	Capture d'écran de la page register	51
5.15	Capture d'écran de la page chatbot	52
5.16	Capture d'écran de la page profile	52

Liste des tableaux

Abstract

Ce mémoire présente le développement de l'application "Safar", une plateforme innovante de tourisme destinée au marché marocain. "Safar" utilise des techniques avancées d'intelligence artificielle pour offrir une expérience personnalisée aux utilisateurs, leur permettant de découvrir et de planifier leurs voyages de manière optimale.

Les principales fonctionnalités de l'application incluent la recommandation de destinations, la création d'itinéraires personnalisés, la réservation d'hébergements et d'activités, ainsi que des informations en temps réel sur les sites touristiques. Le projet aborde également les aspects ergonomiques et graphiques de l'interface utilisateur pour assurer une expérience fluide et intuitive.

L'étude détaille l'analyse des besoins des utilisateurs, la conception de l'application et l'intégration des techniques de traitement du langage naturel et de l'apprentissage automatique pour améliorer l'interaction avec les utilisateurs. En conclusion, "Safar" vise à révolutionner l'industrie du tourisme au Maroc en offrant une solution technologique innovante qui répond aux besoins modernes des voyageurs.

Chapitre 1

Introduction générale

1.1 Contexte

L'évolution rapide des technologies numériques a profondément transformé de nombreux secteurs, y compris celui du tourisme. Avec l'ubiquité des smartphones et l'accessibilité croissante à Internet, les attentes des consommateurs en matière d'expérience utilisateur et de services personnalisés ont considérablement augmenté. Dans ce contexte dynamique, le développement d'applications innovantes joue un rôle crucial pour répondre aux besoins émergents des voyageurs et des prestataires de services touristiques.

1.2 Organisation du mémoire

Ce mémoire est organisé en plusieurs chapitres afin d'aborder de manière systématique les différentes dimensions de notre étude sur l'application "Safarr" dans le domaine du tourisme au Maroc. Voici un aperçu de la structure du mémoire :

- **Chapitre 1 : Introduction générale**

Ce chapitre fournit le contexte général de l'étude et présente l'organisation du mémoire.

- **Chapitre 2 : Analyse des sentiments et revue de la littérature**

Ce chapitre explore les concepts clés de l'analyse des sentiments et examine les recherches antérieures pertinentes.

- **Chapitre 3 : Analyse et Conception**

Ce chapitre détaille l'analyse et la conception de l'application "Safarr", en mettant l'accent sur les aspects fonctionnels, ergonomiques, graphiques, techniques et la modélisation UML.

- **Chapitre 4 : Implémentation et développement**

Ce chapitre discute de l'implémentation pratique de l'application "Safarr", en décrivant les choix technologiques, les défis rencontrés et les solutions mises en œuvre.

- **Chapitre 5 : Évaluation et résultats**

Ce chapitre évalue l'efficacité et les performances de l'application "Safarr", en analysant les retours utilisateurs et en présentant les résultats obtenus.

- Chapitre 6 : Conclusion et perspectives

Ce chapitre résume les principales conclusions de l'étude, propose des recommandations pour l'amélioration continue de l'application, et explore les perspectives futures de développement dans le domaine du tourisme numérique au Maroc.

Chaque chapitre est conçu pour approfondir un aspect spécifique de notre projet, fournissant une vue d'ensemble complète de notre démarche et de nos résultats.

Chapitre 2

Analyse et Conception

Introduction

Ce rapport vise à présenter une analyse détaillée et une conception prospective pour le développement d'une application innovante dans le domaine du tourisme. En répondant aux besoins croissants des utilisateurs modernes, cette initiative vise à combler les lacunes actuelles sur le marché en intégrant des solutions basées sur l'intelligence artificielle et une interface utilisateur intuitive. Ce document examine les défis potentiels, les objectifs stratégiques, et les étapes clés nécessaires pour concrétiser ce projet ambitieux.

2.0.1 Description des besoins fonctionnels

Problématique et périmètre de projet

La problématique centrale de ce projet réside dans la création d'une plateforme robuste qui offre une gamme complète de fonctionnalités pour les voyageurs et les professionnels du tourisme. Le périmètre du projet inclut la réservation simplifiée de services de voyage, la personnalisation des itinéraires en fonction des préférences individuelles, et la fourniture d'informations contextuelles en temps réel. Ces fonctionnalités visent à améliorer l'expérience utilisateur tout en optimisant l'efficacité opérationnelle pour les fournisseurs de services.

Description de la structure générale de l'application

La structure générale de l'application sera conçue pour être modulaire et évolutive, permettant une intégration facile de nouvelles fonctionnalités et une gestion efficace des données. Elle comprendra des modules principaux pour la gestion des utilisateurs, la gestion des réservations, la gestion des contenus dynamiques, et l'intégration avec des APIs externes pour enrichir l'expérience utilisateur.

Fonctionnalités de l'application

Les fonctionnalités principales incluent la recherche et la réservation de vols, d'hébergements et d'activités, la gestion des itinéraires personnalisés, la recommandation de lieux basée sur les préférences utilisateur, ainsi que des alertes en temps réel sur les conditions locales et les événements spéciaux.

2.0.2 Description des besoins ergonomiques

Définir les règles ergonomiques

Les règles ergonomiques définies pour cette application se concentrent sur l’accessibilité, la convivialité et l’efficacité de l’interaction utilisateur. En intégrant des principes de design centré sur l’utilisateur, l’objectif est de minimiser la charge cognitive tout en maximisant la satisfaction utilisateur. L’ergonomie de l’interface utilisateur sera optimisée pour assurer une navigation fluide et intuitive, renforçant ainsi l’attrait et la fidélité des utilisateurs à long terme.

2.0.3 Description des besoins graphiques

L’application adoptera une approche visuelle moderne et attrayante, utilisant des éléments graphiques harmonieux et une palette de couleurs soigneusement sélectionnée pour renforcer l’identité de marque. La conception graphique visera à créer une expérience esthétique et professionnelle tout en assurant une lisibilité et une compréhension claires des informations présentées. Des tests d’utilisabilité seront effectués pour valider l’efficacité des choix visuels et garantir une expérience visuelle agréable pour tous les utilisateurs.

2.0.4 Description des besoins techniques

Outils de développement

Pour le développement de l’application, plusieurs outils et technologies de pointe seront utilisés, incluant des frameworks de développement web et mobile, des services cloud pour le stockage et le traitement des données, ainsi que des APIs pour l’intégration de services tiers. L’architecture technique sera conçue pour être évolutive et sécurisée, tout en répondant aux exigences de performances élevées et de disponibilité constante.

2.0.5 Modélisation UML

Vue Fonctionnelle

La vue fonctionnelle utilisera des diagrammes de cas d’utilisation pour représenter les interactions entre les différents acteurs et les fonctionnalités offertes par l’application. Cela inclura des scénarios de réservation, de personnalisation d’itinéraires, et de gestion de profil utilisateur.

Vue Dynamique

Les diagrammes de séquence seront utilisés pour illustrer le flux d’actions à travers les différentes fonctionnalités de l’application, montrant comment les utilisateurs interagissent avec le système et comment celui-ci répond en conséquence.

Conclusion

En conclusion, ce rapport a identifié les besoins fondamentaux et les défis techniques associés à la conception d’une application innovante dans le secteur du tourisme. En intégrant des solutions technologiques avancées et une conception centrée sur l’utilisateur, ce projet vise à offrir une expérience utilisateur exceptionnelle tout en répondant aux exigences opérationnelles des

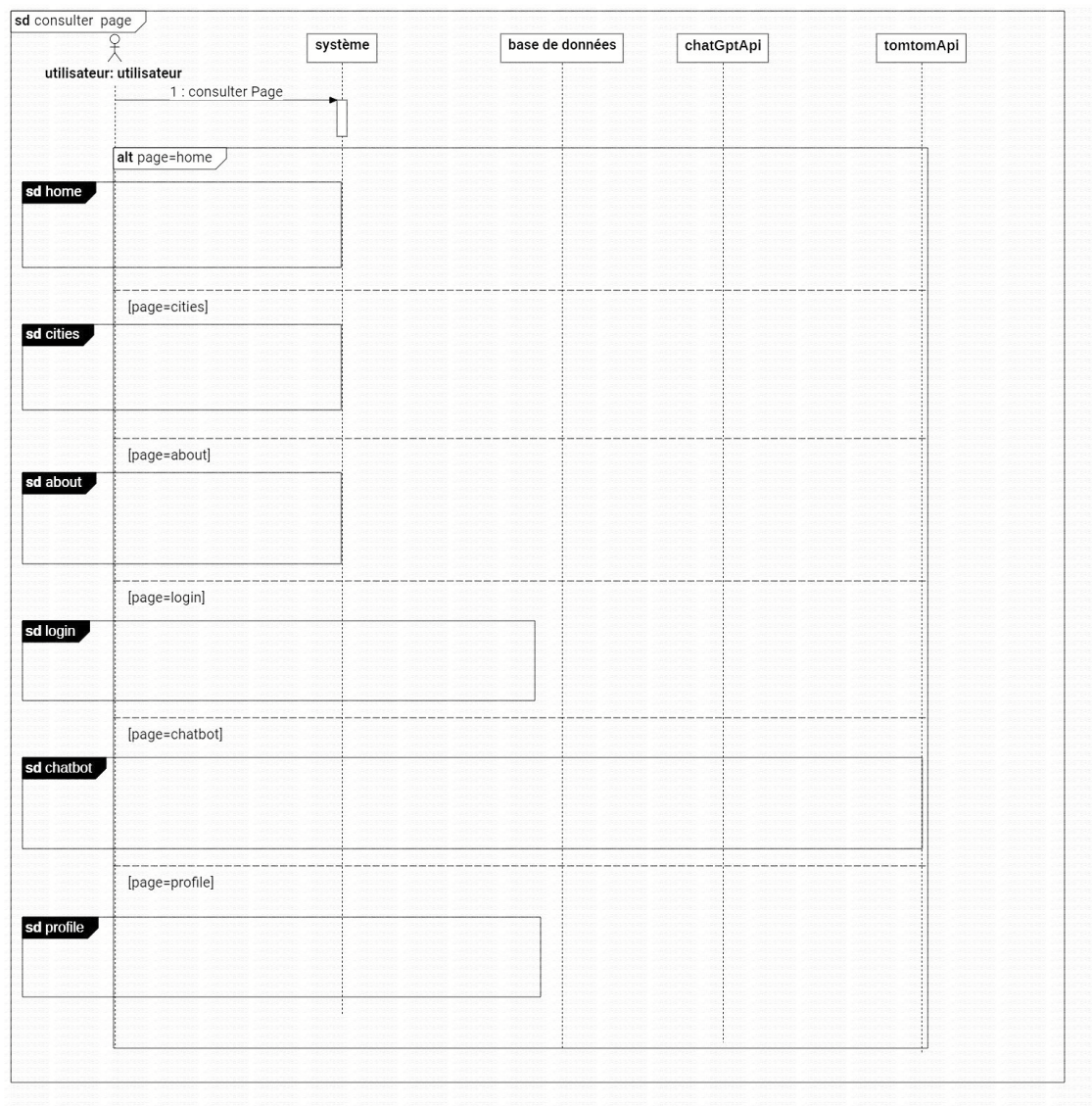


FIGURE 2.2 – Diagramme de séquence du consultation

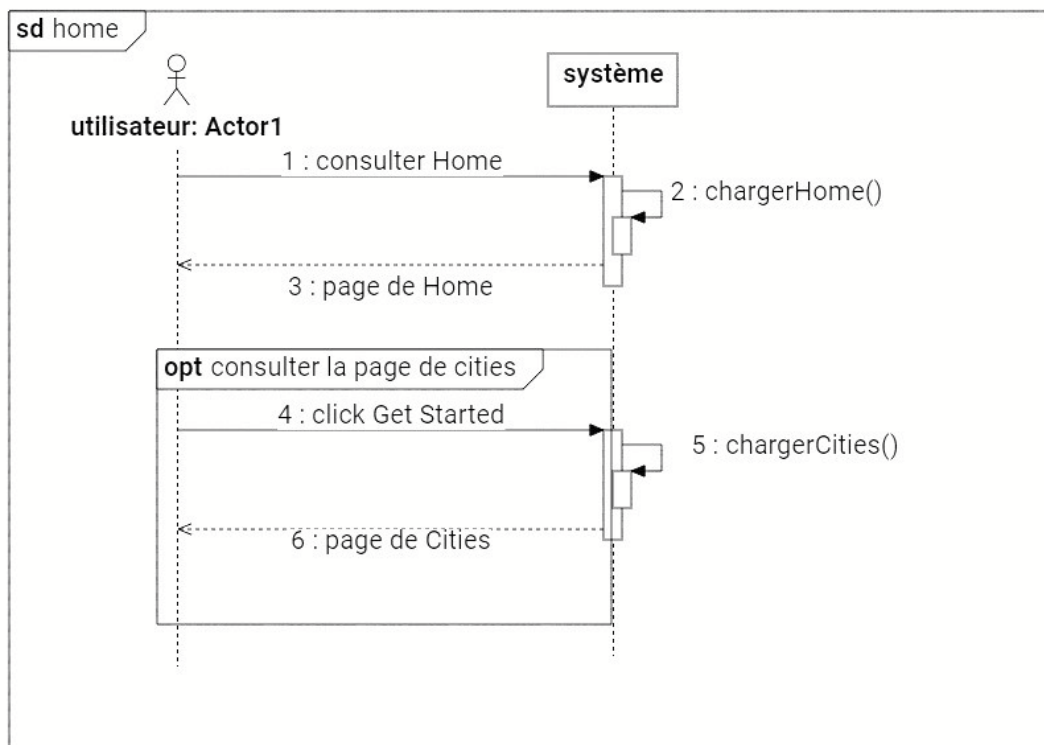


FIGURE 2.3 – Diagramme de séquence de home

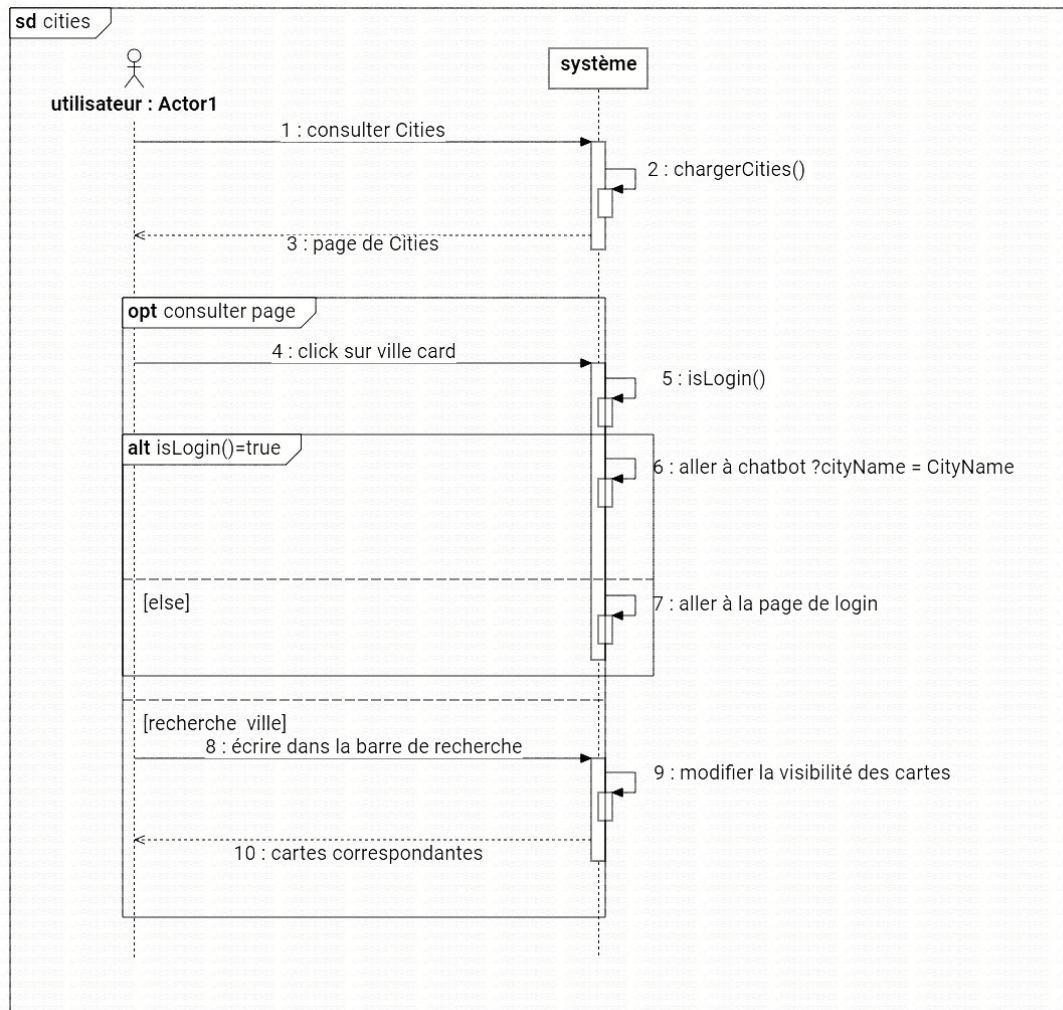


FIGURE 2.4 – Diagramme de séquence de cities

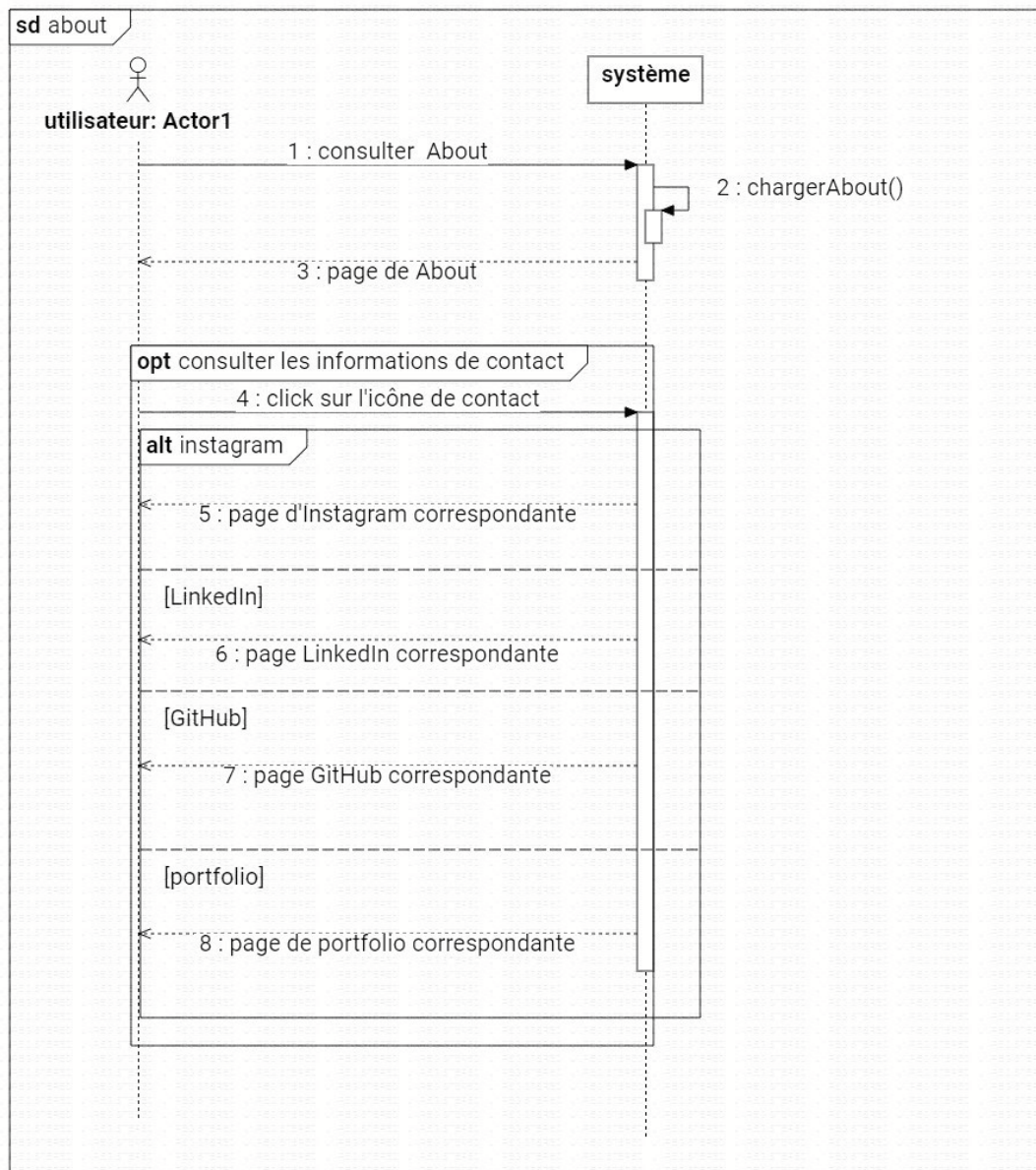


FIGURE 2.5 – Diagramme de séquence d'about

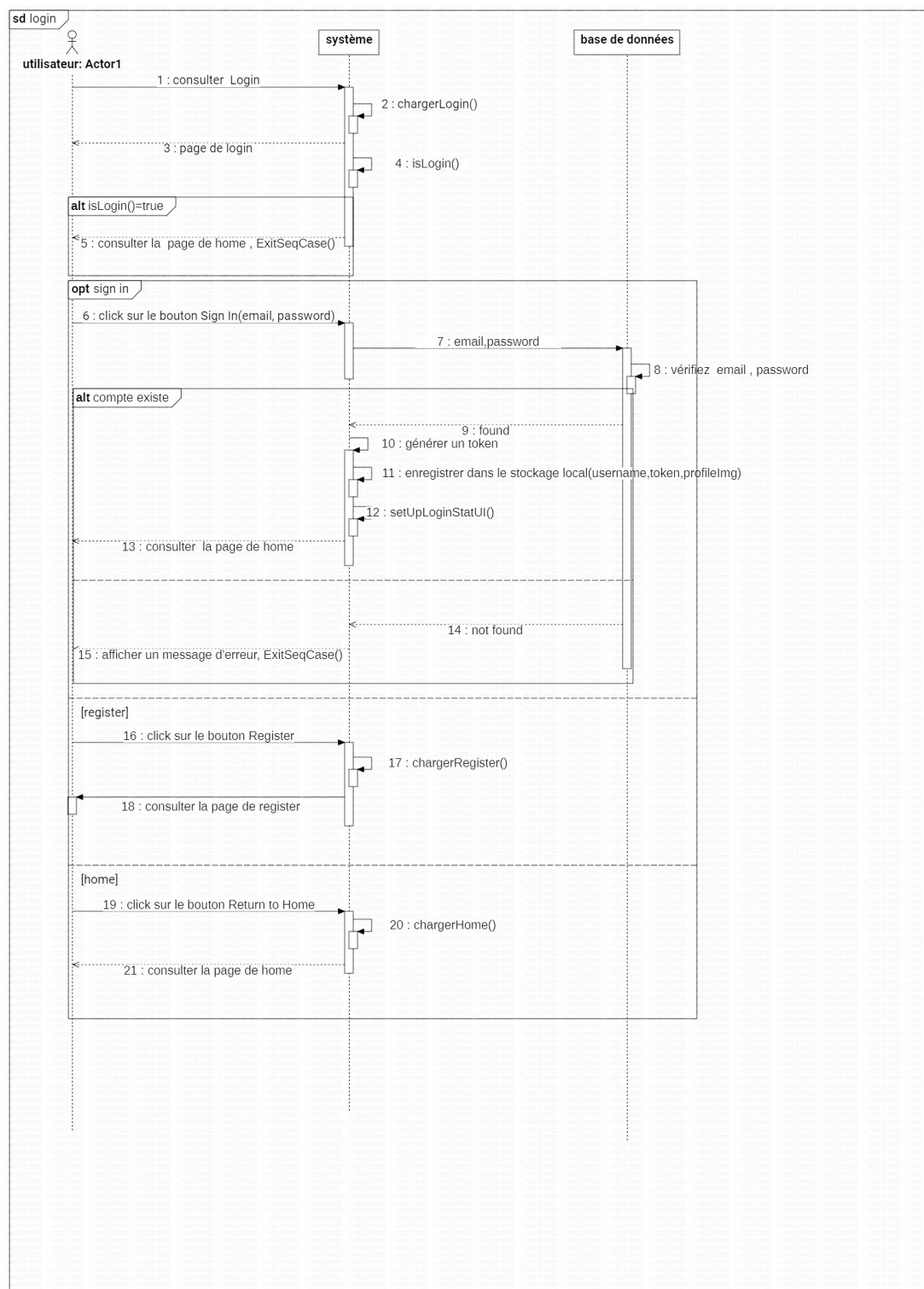


FIGURE 2.6 – Diagramme de séquence du login

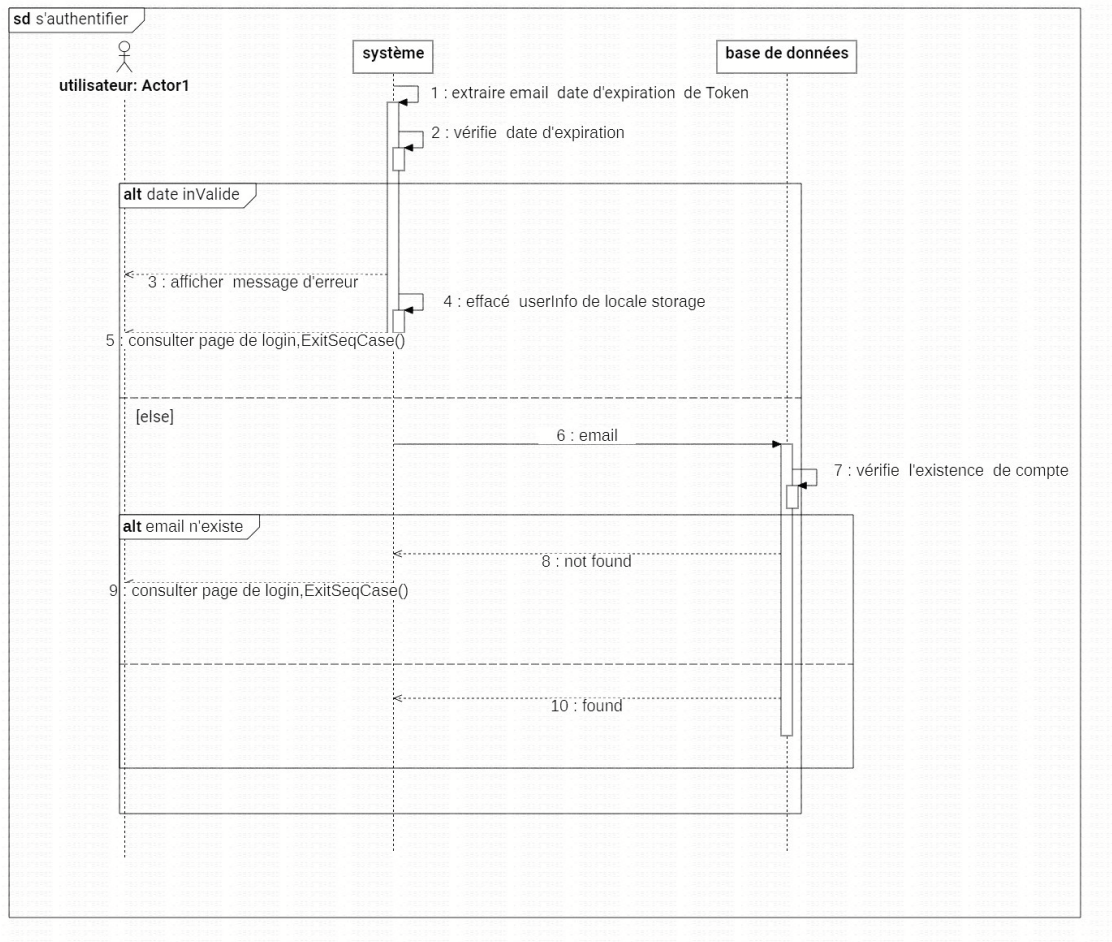


FIGURE 2.7 – Diagramme de séquence d'authentification

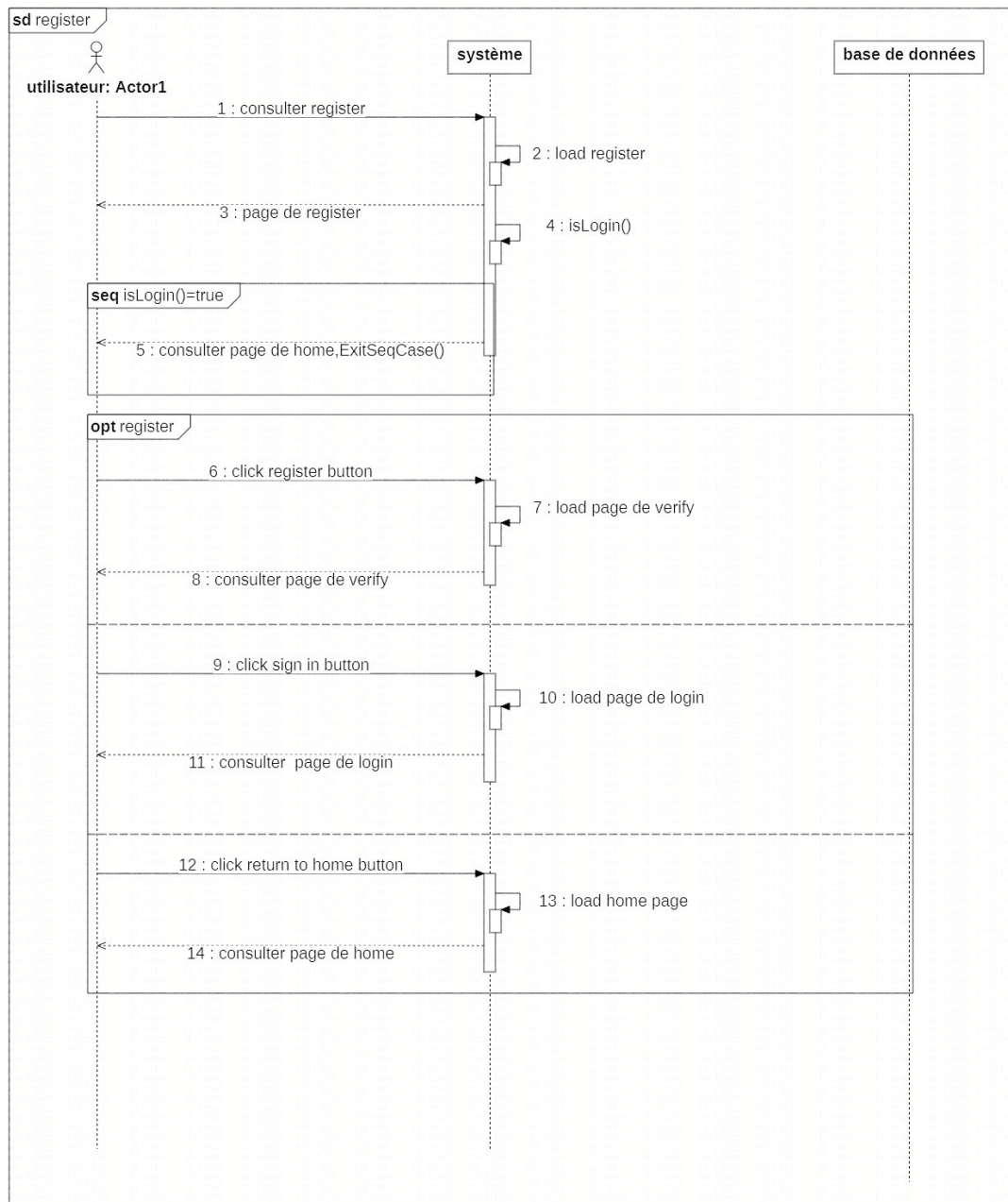


FIGURE 2.8 – Diagramme de séquence du register

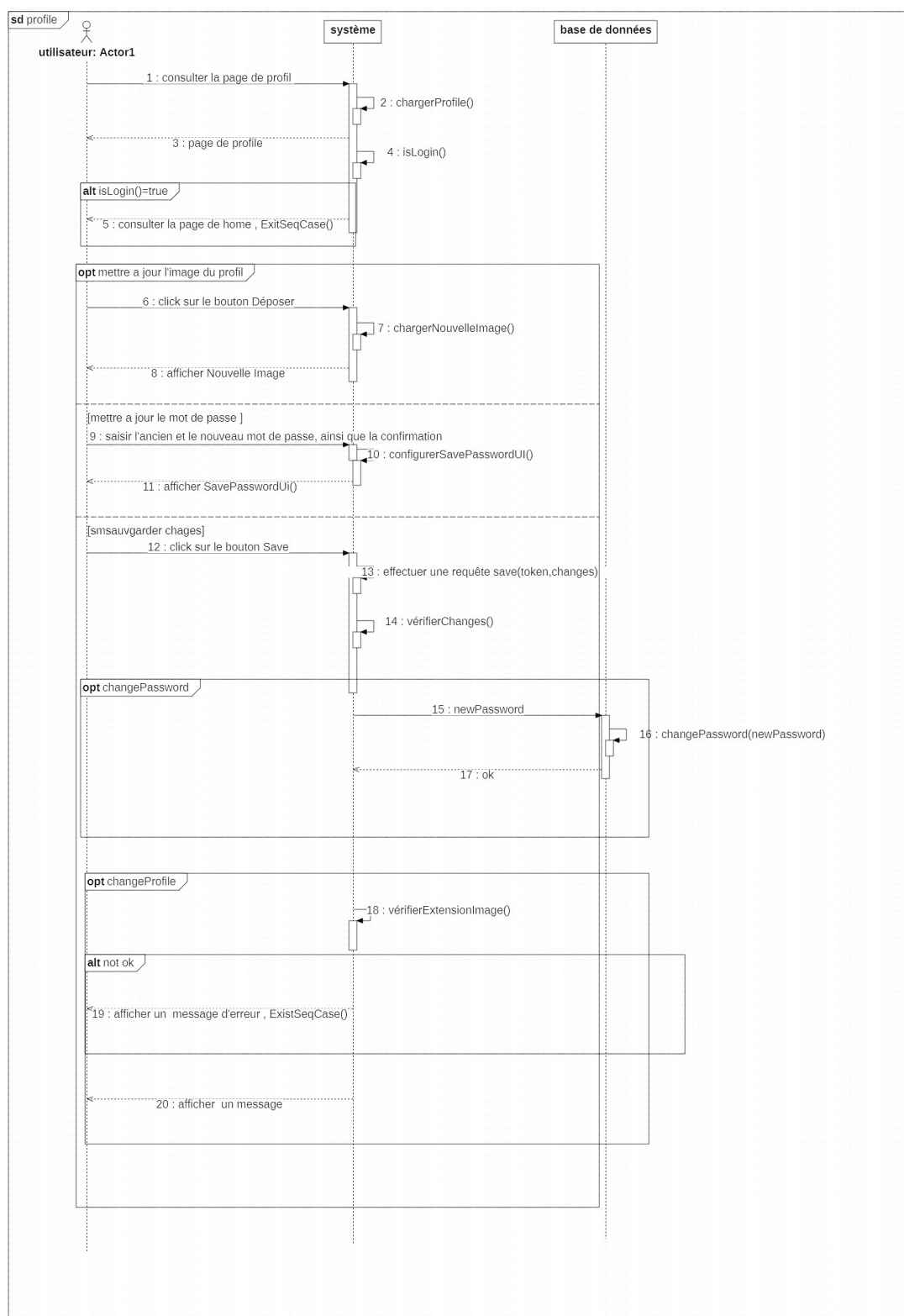


FIGURE 2.10 – Diagramme de séquence du profile

Chapitre 3

Analyse des sentiments et revue de la littérature

Introduction

Ce chapitre explore l'analyse des sentiments, une discipline clé dans le traitement du langage naturel (NLP) qui vise à extraire et à évaluer les opinions, les émotions et les attitudes exprimées dans un texte. En outre, une revue de la littérature sera présentée pour examiner les avancées récentes, les méthodologies et les applications dans ce domaine en évolution rapide.

3.0.1 Sentiments et opinions

Les sentiments et les opinions sont des aspects fondamentaux de l'analyse des sentiments. Les sentiments se réfèrent aux émotions exprimées dans un texte, telles que la joie, la tristesse, la colère, etc., tandis que les opinions représentent les points de vue subjectifs et les jugements sur des sujets spécifiques.

3.0.2 Traitement du langage naturel (NLP)

Le traitement du langage naturel (NLP) est une branche de l'intelligence artificielle qui permet aux machines de comprendre, d'interpréter et de générer un langage humain de manière intelligente. Dans le contexte de l'analyse des sentiments, le NLP joue un rôle crucial en fournissant les outils et les techniques nécessaires pour analyser efficacement le texte.

3.0.3 Analyse des sentiments

Domaines d'applications de l'analyse des sentiments

L'analyse des sentiments trouve des applications dans divers domaines, y compris le marketing, le service client, la politique, la recherche académique, et les médias sociaux. Elle est utilisée pour comprendre les opinions des consommateurs, surveiller la réputation de la marque, analyser les sentiments politiques, et bien plus encore.

Niveaux d'analyse des sentiments

L'analyse des sentiments peut être effectuée à différents niveaux : - **Niveau document** : Évaluer le sentiment global d'un document entier. - **Niveau de phrase** : Analyser le sentiment exprimé dans des phrases individuelles. - **Niveau d'entité** : Évaluer les sentiments associés à des entités spécifiques comme des produits, des entreprises, ou des personnes.

3.0.4 Revue de la littérature

La revue de la littérature examine les recherches antérieures et les publications académiques sur l'analyse des sentiments. Elle explore les méthodologies utilisées, les défis rencontrés, les avancées technologiques, et les applications spécifiques dans divers domaines. Cette section fournira un aperçu critique des travaux existants pour situer le cadre conceptuel et méthodologique de notre étude.

Conclusion

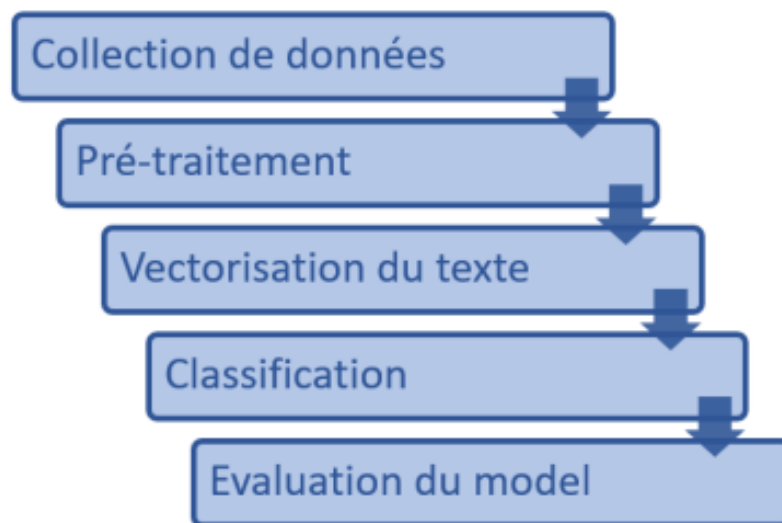
En conclusion, ce chapitre a examiné en profondeur les concepts clés de l'analyse des sentiments, le rôle du traitement du langage naturel (NLP) dans cette discipline, et une revue approfondie de la littérature existante. Il a posé les bases théoriques nécessaires pour aborder notre propre étude sur l'application de l'analyse des sentiments dans le domaine spécifique du tourisme, tel que décrit dans le chapitre précédent sur l'Analyse et Conception.

Chapitre 4

Procédure d'analyse des sentiments

Introduction

L'analyse des sentiments est une méthode essentielle pour extraire et comprendre les opinions et les émotions exprimées dans les textes. Ce chapitre explore les étapes clés de cette procédure.



4.0.1 Collection de données

La collecte de données implique l'agrégation de textes à partir de diverses sources telles que les réseaux sociaux, les critiques de produits et les forums en ligne.

4.0.2 Prétraitement de la base de données

Tokenization

La tokenisation divise les phrases en tokens ou mots individuels, facilitant ainsi l'analyse ultérieure en isolant chaque unité significative de texte.

Suppression des mots vides

Les mots vides (ou stopwords) sont des mots courants qui ne contribuent pas significativement au sens d'un texte. Leur suppression aide à réduire le bruit et à améliorer la précision de l'analyse.

Stemming

Le stemming réduit les mots à leur racine, ce qui permet de normaliser différentes variantes d'un même mot et de regrouper les termes similaires lors de l'analyse.

4.0.3 Word Embedding (Vectorisation du texte)

Term Frequency - Inverse Document Frequency (TF-IDF)

TF-IDF assigne des poids aux mots en fonction de leur fréquence dans un document par rapport à leur fréquence dans le corpus entier, mettant en évidence les mots les plus importants pour chaque texte.

Le poids TF-IDF ($W_{t,Doc}$) d'un terme (t) dans un document (Doc) est calculé selon la formule suivante :

$$W_{t,Doc} = TF_{t,Doc} \cdot \log \frac{N}{N_t}$$

FIGURE 4.1 – $W_{t,Doc} = TF_{t,Doc} * \log(N/N_t)$

Où :

- $TF_{t,Doc}$: Fréquence du terme (t) dans le document (Doc).
- N : Nombre total de documents dans le corpus.
- N_t : Nombre de documents contenant le terme (t).
- Le principal inconvénient de cette méthode est qu'elle considère le document comme un ensemble de mots indépendants les uns des autres, sans prendre en compte la structure ou le contexte dans lequel les mots apparaissent

Word2Vec

Word2vec est une technique populaire en traitement du langage naturel (NLP) pour apprendre des représentations vectorielles de mots à l'aide de réseaux de neurones peu profonds. Développée par Tomas Mikolov en 2013 chez Google, elle attribue à chaque mot un vecteur numérique qui capture sa signification.

L'interprétation des vecteurs de mots se fait en les comparant entre eux et en utilisant des équations simples. Par exemple, "Directeur" - "homme" + "femme" = "Directrice", ce qui démontre la capacité de ces vecteurs à capturer la sémantique des mots.

La force de Word2vec réside dans l'apprentissage des vecteurs à partir du contexte d'apparition des mots. Cela permet d'obtenir des vecteurs où les mots au sens similaire partagent des représentations numériques proches. Un inconvénient de Word2vec est que chaque mot n'a qu'un seul vecteur. Or, dans la réalité, le sens d'un mot peut varier en fonction du contexte.

Deux architectures principales : C-BOW et Skip-gram :

Word2vec se décline en deux architectures principales :

- **C-BOW (Continuous Bag Of Words)** : Elle prend le contexte d'un mot en entrée et tente de prédire le mot cible.

- **Skip-gram** : Fonctionnement inverse de C-BOW, elle prédit le contexte à partir du mot cible.

Choix entre C-BOW et Skip-gram :

C-BOW est généralement plus rapide à entraîner et donc adapté aux petits corpus, tandis que Skip-gram est plus efficace pour traiter de grands corpus.

FastText

FastText est une méthode de vectorisation qui traite le texte avant qu'il ne soit analysé par des algorithmes d'apprentissage automatique. Contrairement à d'autres techniques, FastText prend en compte la structure interne des mots plutôt que de se concentrer uniquement sur les représentations des mots eux-mêmes.

Cette méthode est particulièrement utile pour les langues avec une richesse morphologique, car elle permet d'apprendre des représentations pour différentes formes morphologiques de mots de manière indépendante. FastText se base sur la technique des n-grammes, qui divise les mots en parties de même dimension.

Exemple :

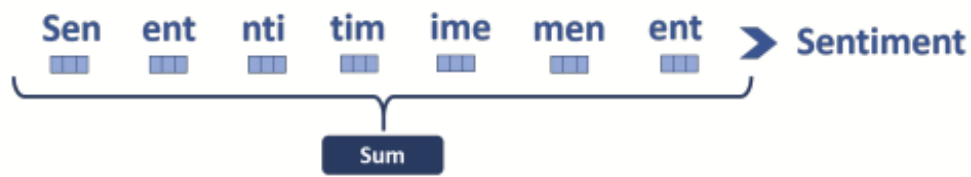


FIGURE 4.2 – FastText

GloVe

GloVe (Global Vector for Word Representation) est une nouvelle approche d'apprentissage non supervisé pour générer des représentations vectorielles de mots. Cette technique se distingue par sa rapidité, sa simplicité et sa performance supérieure aux méthodes existantes.

Principe de fonctionnement :

GloVe fonctionne en analysant les co-occurrences de mots dans un corpus. Il construit une matrice de co-occurrence où chaque élément représente la fréquence de co-occurrence de deux mots distincts. Ensuite, le modèle applique une régression log-bilinéaire pour apprendre des vecteurs vectoriels pour chaque mot.

Avantages de GloVe :

- Rapidité et simplicité : GloVe est plus rapide et plus simple à implémenter que d'autres méthodes de représentation de mots, comme Word2Vec.
- Meilleure performance : GloVe a montré des performances supérieures à Word2Vec dans diverses tâches de traitement du langage naturel, telles que l'analogie de mots et la sémantique des phrases.

Téléchargement :

Le modèle GloVe est disponible en téléchargement gratuit sur son site officiel [1].

Résolution du problème de Word2Vec :

GloVe résout un problème clé de Word2Vec, qui est sa sensibilité à la fréquence des mots. Word2Vec tend à attribuer des vecteurs plus puissants aux mots plus fréquents, ce qui peut biaiser les résultats dans certaines tâches. GloVe utilise une approche pondérée pour atténuer cet effet, ce qui lui permet d'obtenir des représentations de mots plus précises et plus robustes.

En résumé, GloVe est une technique prometteuse pour la représentation de mots qui offre une combinaison unique de rapidité, de simplicité et de performance supérieure.

BERT

BERT (Bidirectional Encoder Representations from Transformers) est un modèle d'embedding. Il utilise un réseau de neurones pour représenter du texte, qu'il s'agisse de mots ou de phrases, sous forme de vecteur numérique en fonction de son contexte.

-**Bidirectionnel** : BERT peut travailler dans les deux sens d'une phrase, ce qui lui permet de capturer des informations contextuelles plus riches.

-**Représentation de l'encodeur** : Il s'agit de la base des modèles de traitement automatique du langage naturel (NLP). Cette représentation permet de simplifier les données textuelles pour l'ordinateur tout en permettant des modifications généralisées.

L'architecture de BERT repose sur les transformeurs. Actuellement, deux variantes sont disponibles :

- **BERT Base** : Comprend 12 couches (blocs de transformateurs), 12 têtes d'attention et 110 millions de paramètres.

- **BERT Large** : Comprend 24 couches (blocs de transformateurs), 16 têtes d'attention et 340 millions de paramètres.

Comparaison générale

Les modèles Word2Vec génèrent des embeddings qui sont indépendants du contexte : chaque mot a une seule représentation vectorielle pour tous les contextes d'utilisation.

FastText, quant à lui, utilise une approche basée sur les n-grammes, ce qui lui permet de capturer partiellement le contexte. Cependant, il est moins performant que Word2Vec en termes de représentation contextuelle.

En revanche, le modèle BERT surpasse à la fois Word2Vec et FastText en termes de représentation contextuelle. BERT génère des embeddings qui dépendent du contexte, ce qui signifie qu'un même mot peut avoir plusieurs représentations vectorielles en fonction de son utilisation dans différents contextes. Cette capacité à capturer le contexte rend les embeddings BERT particulièrement adaptés pour des tâches de traitement du langage naturel nécessitant une compréhension contextuelle plus fine

4.0.4 Classification

La phase de classification est l'étape principale de l'analyse des sentiments, nécessitant l'exécution de toutes les tâches précédentes. La diversité des choix parmi les différents algorithmes, basés sur des méthodes classiques (basées sur des règles), d'apprentissage automatique et d'apprentissage profond, fait de cette étape un défi fondamental qui met en lumière les différences et les qualités de chaque modèle.

Les techniques d'analyse des sentiments sont illustrées dans la figure suivante :

Méthodes classiques

Dans un premier temps, nous trouvons l'approche lexicale, qui repose sur l'utilisation d'un dictionnaire des sentiments contenant des mots d'opinion pour déterminer la polarité des données.

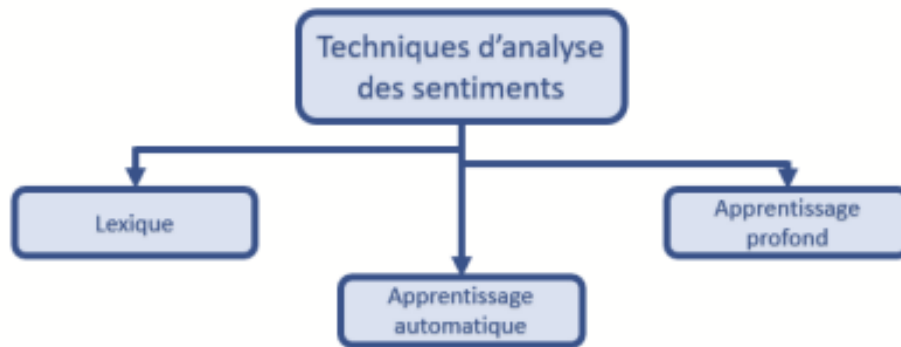


FIGURE 4.3 – Techniques d’analyse des sentiments

Cette approche attribue des scores de sentiment aux mots d’opinion, indiquant s’ils sont positifs, négatifs ou neutres.

Cette approche comprend deux types, basés sur :

- **Dictionnaire** : Dans cette approche, un ensemble de mots d’opinion est collecté manuellement pour former une liste de départ. Ensuite, des dictionnaires sont recherchés pour trouver des synonymes et des antonymes des mots. Les synonymes trouvés sont ajoutés à la liste de départ, et ce processus se répète jusqu’à ce qu’aucun nouveau mot ne soit trouvé. Cependant, cela peut présenter des difficultés pour contextualiser ou orienter les mots dans un domaine spécifique.

- **Corpus** : Un corpus est une collection d’écrits, souvent sur un sujet spécifique. Dans cette approche, la liste de départ est préparée et étendue en utilisant le texte du corpus. Cette méthode peut être réalisée de deux manières :

- **Approche statistique** : Cette méthode utilise les mots de cooccurrence dans le corpus pour déterminer la polarité. Par exemple, si un mot apparaît principalement dans un texte positif, il est considéré comme positif.

- **Approche sémantique** : Cette méthode calcule les valeurs de sentiment en utilisant le principe de similarité entre les mots. WordNet peut être utilisé pour trouver les synonymes et les antonymes d’un mot donné, puis calculer la valeur du sentiment.

Méthodes Machine Learning

Méthodes supervisées

Naive Bayes Naive Bayes est un classificateur probabiliste simple basé sur le théorème de Bayes, souvent utilisé pour la classification de texte en fonction des probabilités conditionnelles.

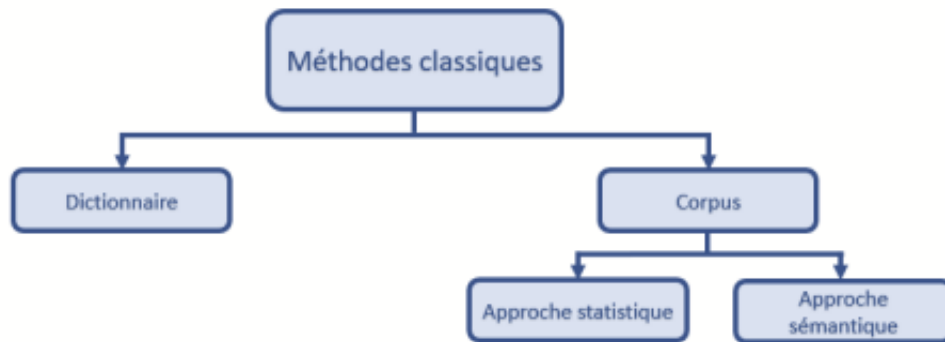


FIGURE 4.4 – Récapitulatif des méthodes classiques d’analyse des sentiments

Bayesian Network Les réseaux bayésiens utilisent des modèles graphiques pour représenter les dépendances probabilistes entre les variables, adaptés à la modélisation de l’incertitude dans les données de texte.

Maximum Entropy La méthode du maximum d’entropie attribue des poids aux caractéristiques du texte en maximisant l’entropie conditionnelle sous contraintes, adaptée à la classification de texte basée sur des caractéristiques complexes.

Support Vector Machine Les machines à vecteurs de support utilisent des hyperplans pour séparer les classes de texte dans un espace multidimensionnel, efficaces pour la classification de texte avec des marges maximales.

Decision Tree Les arbres de décision structurent les décisions en une série de questions hiérarchiques basées sur les caractéristiques du texte, facilitant l’interprétation des règles de classification.

Méthodes Deep Learning

Réseaux de neurones L’idée clé des réseaux de neurones est d’extraire des entités à partir d’une combinaison linéaire des données d’entrée, puis de modéliser la sortie en tant que fonction non linéaire de ces entités. Ce type d’algorithme est apparu ces dernières années comme un domaine d’opportunités inhabituelles pour la recherche dans le domaine de l’analyse des sentiments.

La couche d’entrée est la première étape d’un réseau de neurones. Elle prend comme entrée les données qui doivent être au format numérique, ce qui nécessite souvent un prétraitement des données pour les convertir en nombres.

La/les couche(s) cachée(s) est/sont le cœur d’un réseau de neurones, car c’est à cette étape que le traitement des données se fait. Cette couche est composée de nœuds, chacun essayant d’appliquer une combinaison des données d’entrée avec un ensemble de coefficients ou poids, qui servent généralement à amplifier ou atténuer cette entrée et à générer des produits.

Ces produits de pondération d’entrée sont ensuite additionnés, puis la somme est transmise à la fonction d’activation d’un nœud pour déterminer si et dans quelle mesure ce signal doit

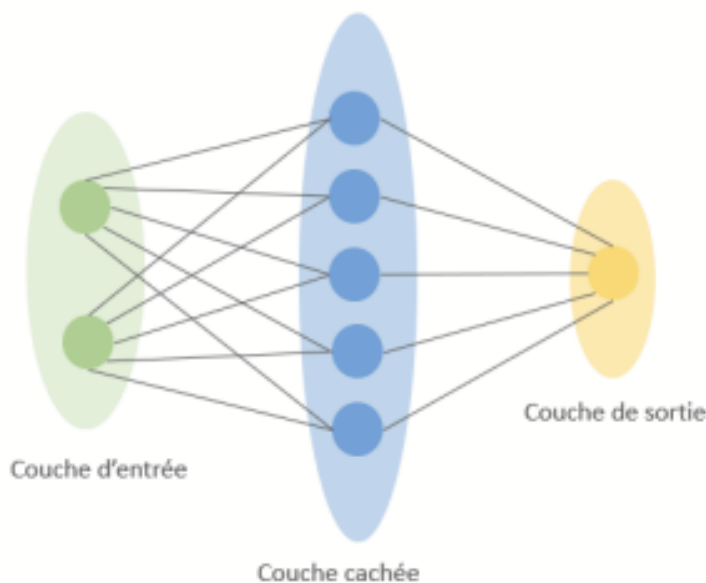


FIGURE 4.5 – Structure générale d'un réseau de neurones

progresser davantage dans le réseau pour affecter le résultat final.

Réseaux de neurones convolutifs Les réseaux de neurones convolutifs appliquent des filtres convolutifs pour extraire des caractéristiques locales dans les textes, adaptés à la classification de texte basée sur la structure spatiale.

Réseaux de neurones récurrents Les réseaux de neurones convolutifs, ou CNN (Convolutional Neural Networks), sont spécialement conçus pour analyser finement des données exprimées sous forme de grille, typiquement les pixels d'une image. Ils se sont avérés efficaces dans l'analyse de texte et le traitement d'images.

Un réseau CNN est construit en enchaînant une couche de convolution et une couche de pooling, puis se termine par une couche de neurones totalement connectés. L'architecture d'un réseau CNN pour une tâche de classification de texte est présentée dans la figure ci-dessous.

La figure précédente illustre l'architecture d'un réseau de neurones convolutifs pour un exemple sous forme de phrase en anglais : "wait for the video and don't rent it".

Dans une convolution de matrice, deux matrices de dimensions incompatibles sont multipliées. Cette opération consiste généralement à appliquer un filtre ou un "kernel" sur la grille ou la matrice de données. Le kernel est une matrice de dimensions inférieures à celles de la matrice de données. On positionne le centre du noyau sur l'élément de la matrice que l'on souhaite modifier, puis on applique un produit scalaire entre le recouvrement par le kernel sur la matrice

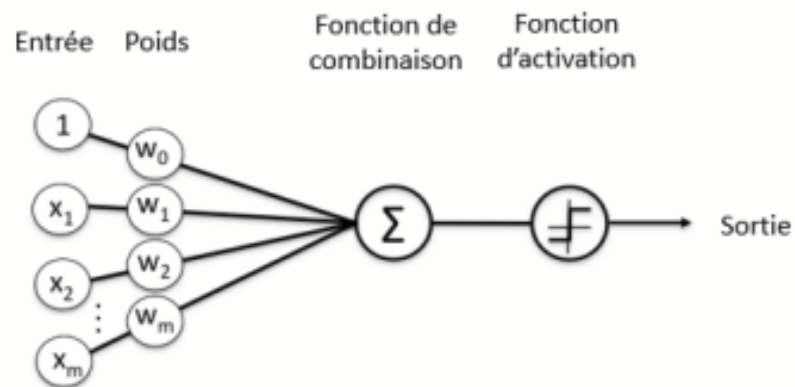


FIGURE 4.6 – Fonctionnement d'un réseau de neurones

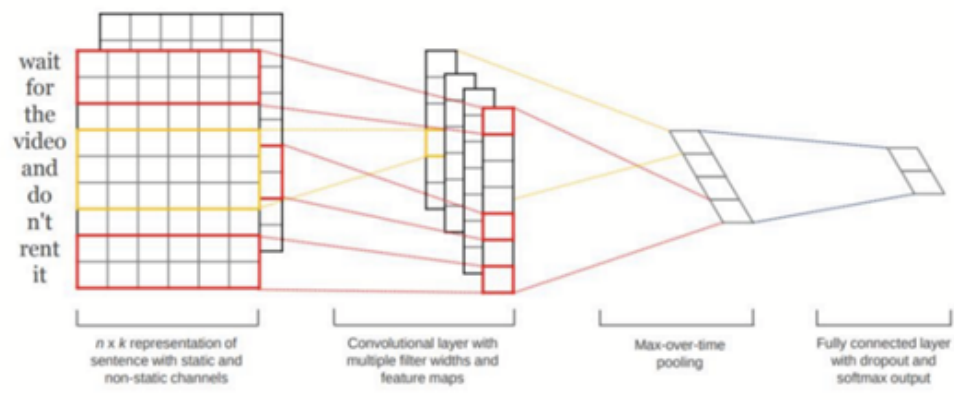


FIGURE 4.7 – Architecture d'un réseau de neurones convolutif

et le kernel lui-même. Ensuite, on déplace le noyau d'un pas sur l'élément suivant et ainsi de suite. Cette opération permet d'extraire des caractéristiques spécifiques pour chaque filtre choisi, produisant ainsi une carte de caractéristiques.

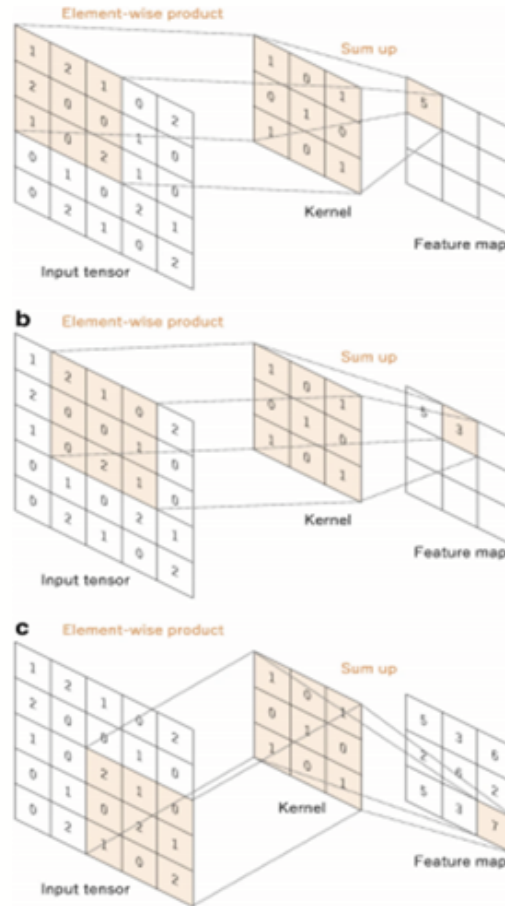


FIGURE 4.8 – Convolution d'une matrice via un filtre

Pour réduire la dimension de la grille obtenue après la convolution, on applique une opération de pooling. Cette opération consiste à faire glisser une petite fenêtre pas à pas sur toutes les parties de la grille et à prendre la valeur maximum, moyenne ou minimum de cette fenêtre à chaque pas. Ainsi, la grille obtenue est de dimension réduite par rapport à la grille initiale, permettant au réseau de détecter des informations importantes sans se soucier de leur position précise.

Le sous-échantillonnage est nécessaire pour réduire la quantité d'informations générées par cette opération. Le Pooling, une méthode qui réduit la taille d'une grande grille d'informations tout en préservant les informations les plus importantes, est souvent utilisé à cette fin

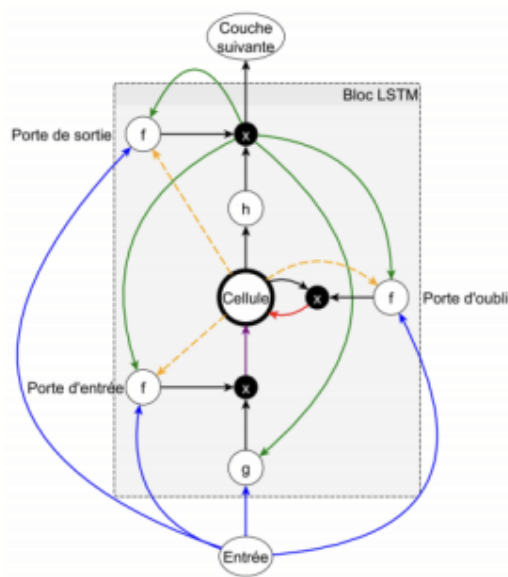


FIGURE 4.9 – Architecture d'un seul neurone LSTM

Mémoire longue à court terme (LSTM) Les réseaux de mémoire à long terme et court terme (LSTM) sont une extension des réseaux de neurones récurrents (RNN) qui étendent essentiellement la capacité de mémorisation. Ils sont bien adaptés pour tirer des leçons d'expériences importantes qui présentent de très longs décalages temporels.

Les unités d'un LSTM sont utilisées comme unités de construction pour les couches d'un RNN, souvent appelé réseau LSTM.

Cette mémoire peut être considérée comme une cellule fermée, ce qui signifie que la cellule décide de stocker ou de supprimer des informations en fonction de l'importance qu'elle leur accorde. L'importance est attribuée par le biais de poids, qui sont également appris par l'algorithme. Ainsi, le réseau apprend au fil du temps quelles informations sont importantes et lesquelles ne le sont pas.

La cellule mémoire peut être contrôlée par trois portes comme des vannes :

- La porte d'entrée décide si l'entrée doit modifier le contenu de la cellule.
- La porte d'oubli décide s'il faut remettre à zéro le contenu de la cellule.
- La porte de sortie décide si le contenu de la cellule doit influencer la sortie du neurone..

Conclusion

En conclusion, ce chapitre a exploré en détail la procédure d'analyse des sentiments, en mettant l'accent sur les étapes de collecte de données, de prétraitement, de vectorisation du texte et de classification. Les méthodes présentées offrent un cadre méthodologique robuste pour

l'analyse efficace des opinions et des émotions exprimées dans les textes, adaptées à diverses applications dans le domaine du traitement du langage naturel.

Chapitre 5

Interfaces de l'application

Introduction

Ce chapitre explore les interfaces de l'application développée, en mettant l'accent sur les outils utilisés, le prétraitement des données, le word embedding, la classification et les différentes interfaces utilisateur.

5.0.1 Les outils utilisés

Les outils utilisés pour développer les interfaces :

Conception et Design :

- StarUML
- Figma

Environnement :

- Microsoft Visual Studio Code
- Microsoft Windows 11 (DEVs Computer)
- Linux Ubuntu Container (Servers)

Deployment et Tests :

- Git
- GitHub
- RocmineLab Servers
- AWS Amazon

- Live Server

Front end :

- HTML5

- CSS3

- JavaScript

Backend et Base de données :

- Flask

- MongoDB

API Endpoints :

- OpenAI

- Google Maps

5.0.2 Prétraitement de données (captures d'écran)

Suppression des URLs, émojis, mots vides et ponctuations

Le prétraitement des données inclut des étapes pour nettoyer les textes en supprimant les URLs, les émojis, les mots vides et la ponctuation, assurant des données propres pour l'analyse ultérieure.

Tokenization

La tokenization divise les textes en tokens ou mots individuels, facilitant ainsi l'analyse et le traitement automatisé des données textuelles.

5.0.3 Word Embedding (captures d'écran)

Word2Vec

Word2Vec génère des représentations vectorielles de mots à partir de grands corpus textuels, capturant les relations sémantiques entre les mots.

Output :

love product amazing', 'movie terrible hated', 'weather nice today', 'feeling sad'

FastText

FastText étend Word2Vec en incorporant des informations de sous-mots, améliorant la représentation des mots et des expressions complexes.

GloVe

GloVe utilise des matrices de co-occurrence pour apprendre des vecteurs de mots, offrant une représentation pondérée des mots basée sur leurs relations dans le corpus.

5.0.4 Classification (captures d'écran)

LR

La régression logistique (LR) est utilisée pour la classification binaire et multiclasse des textes, basée sur des fonctions logistiques pour estimer les probabilités de classe.

Naive Bayes

Naive Bayes utilise le théorème de Bayes pour la classification probabiliste des textes, considérant l'indépendance conditionnelle des caractéristiques pour prédire les classes.

DT

Les arbres de décision (DT) structurent les décisions à partir des caractéristiques des textes, formant des règles de classification hiérarchiques pour la prise de décision.

SVM

Les machines à vecteurs de support (SVM) définissent des hyperplans pour séparer les classes de textes dans un espace multidimensionnel, adaptées à la classification avec marges maximales.

5.0.5 Interfaces (captures d'écran)

Page d'accueil

Capture d'écran de la page d'accueil de l'application, montrant les fonctionnalités principales et la navigation utilisateur.

Page Cities

Capture d'écran de la page cities qui cite les différentes villes marocaines.

Page About

Capture d'écran de la page about qui donne des informations sur la plateforme SAFAR et sur l'équipe du SAFAR.

Page Login

Capture d'écran de la page login pour que l'utilisateur se connecte à la plateforme.

Page Register

Capture d'écran de la page login pour que l'utilisateur s'enregistre à la plateforme.

Page Chatbot

Capture d'écran de la page chatbot où l'utilisateur peut communiquer et utiliser l'intelligence artificielle de la plateforme SAFAR.

Page profile

Capture d'écran de la page profile où l'utilisateur peut modifier les informations de son compte.

Conclusion

En conclusion, ce chapitre a exploré les interfaces de l'application, détaillant les outils utilisés, les étapes de prétraitement des données, les méthodes de word embedding, de classification, et présentant les interfaces utilisateur principales. Ces aspects sont essentiels pour assurer une expérience utilisateur fluide et une analyse précise des textes dans l'application développée.

```

1
2 -----Second DataSet Format :
3
4                                new_reviews  score
5  0    our stay at hotel allegro was amazing the staf...    1
6  1    wonderful clean modern rooms with huge comfort...   -1
7  2    spent quick days and was impressed by both aga...    1
8  3    we visited for one week in december and had su...    1
9  4    good hotel in very good location with all good...    1
10 ...                                ...    ...
11 7846 comfortable and homelybeen to more luxurious h...    1
12 7847 rd visit had wonderful rd visit was there for ...    1
13 7848 ideal winter sun breakwe stayed in this hotel ...    1
14 7849 delightful placestayed here in april and had a...    1
15 7850 nice placethe hotel is good expected more trad...    1
16
17 [7851 rows x 2 columns]
18 (7851, 2)
19 :
20
21 -----Second DataSet Format :
22
23                                Review  Rating
24  0    nice hotel expensive parking got good deal sta...    4
25  1    ok nothing special charge diamond member hilt...    2
26  2    nice rooms not 4* experience hotel monaco seat...    3
27  3    unique, great stay, wonderful time hotel monac...    5
28  4    great stay great stay, went seahawk game aweso...    5
29 ...                                ...    ...
30 20486 best kept secret 3rd time staying charm, not 5...    5
31 20487 great location price view hotel great quick pl...    4
32 20488 ok just looks nice modern outside, desk staff ...    2
33 20489 hotel theft ruined vacation hotel opened sept ...    1
34 20490 people talking, ca n't believe excellent ratin...    2
35
36 [20491 rows x 2 columns]
37 (20491, 2)

```

FIGURE 5.1 – Exemple de prétraitement des données

```

1  Evaluation Report LR:
2      precision    recall  f1-score   support
3
4      +         0.91      0.96      0.93       4186
5      -         0.86      0.74      0.80       1483
6
7      accuracy                0.90       5669
8      macro avg         0.89      0.85      0.87       5669
9      weighted avg        0.90      0.90      0.90       5669
10
11 Accuracy Score LR: 0.9015699417886752
12

```

FIGURE 5.2 – Exemple de classification avec régression logistique

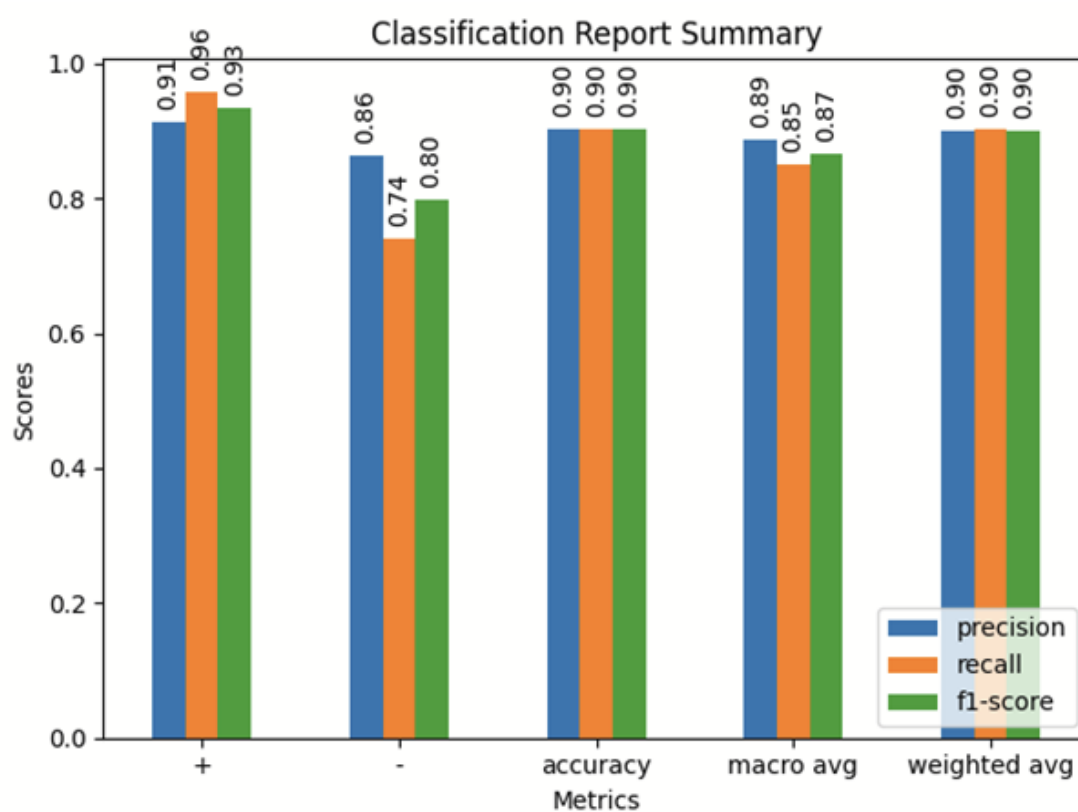


FIGURE 5.3 – Rapport Résultat du model LR

1	Evaluation Report NB:				
2		precision	recall	f1-score	support
3					
4	+	0.75	1.00	0.86	4186
5	-	0.99	0.08	0.14	1483
6					
7	accuracy			0.76	5669
8	macro avg	0.87	0.54	0.50	5669
9	weighted avg	0.82	0.76	0.67	5669
10					
11	Accuracy Score NB: 0.757982007408714				
12					

FIGURE 5.4 – Exemple de classification avec Naive Bayes

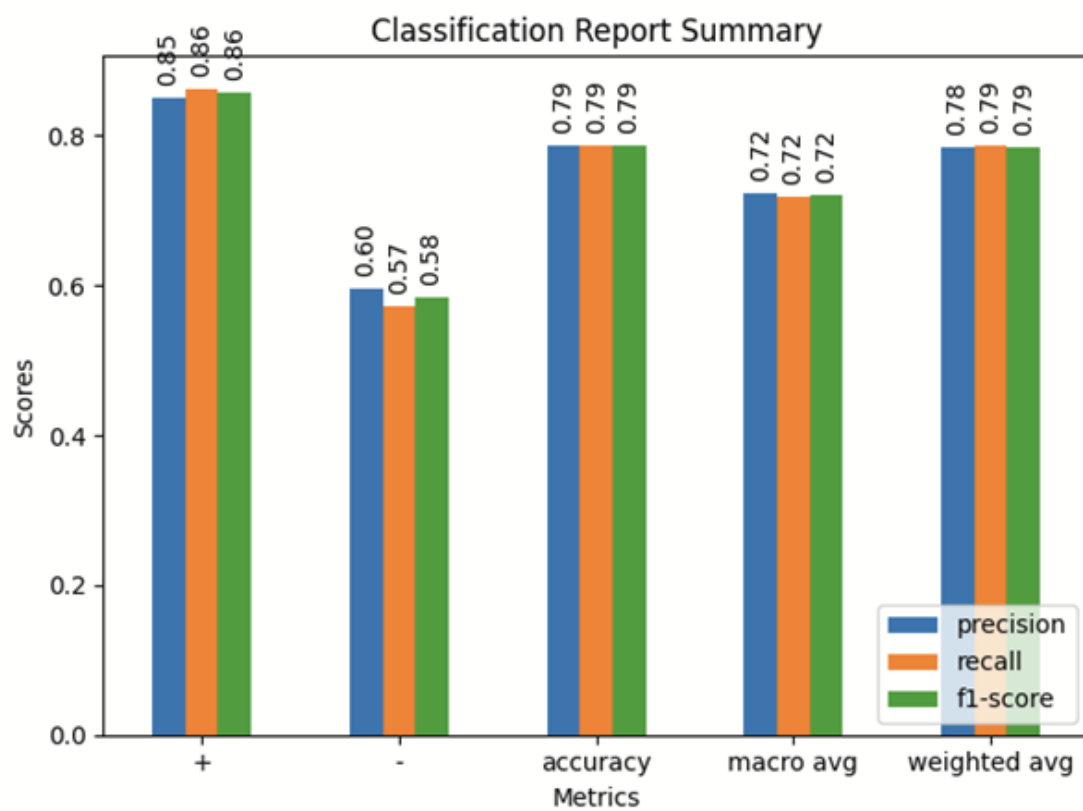


FIGURE 5.5 – Rapport Résultat du model Naive Bayes

```
1  Evaluation Report DT:
2      precision    recall  f1-score   support
3
4      +         0.85      0.86      0.86      4186
5      -         0.60      0.57      0.58      1483
6
7      accuracy                0.79      5669
8      macro avg              0.72      0.72      0.72      5669
9      weighted avg           0.78      0.79      0.79      5669
10
11  Accuracy Score DT: 0.786734873875463
12
```

FIGURE 5.6 – Exemple de classification avec arbres de décision

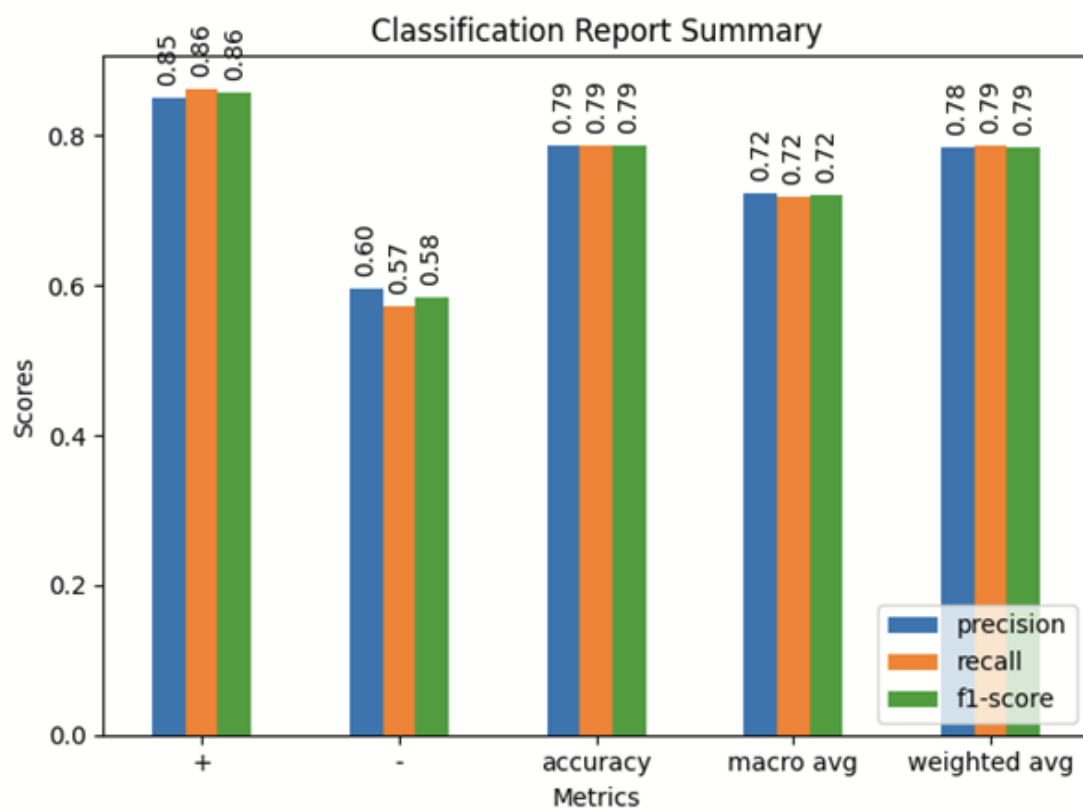


FIGURE 5.7 – Rapport Résultat du model arbres de décision DT

```
1  Evaluation Report SVM:
2      precision    recall  f1-score   support
3
4      +         0.91      0.96      0.94      4186
5      -         0.87      0.74      0.80      1483
6
7      accuracy                0.90      5669
8      macro avg         0.89      0.85      0.87      5669
9      weighted avg         0.90      0.90      0.90      5669
10
11 Accuracy Score SVM: 0.9033339213265126
12
```

FIGURE 5.8 – Exemple de classification avec SVM

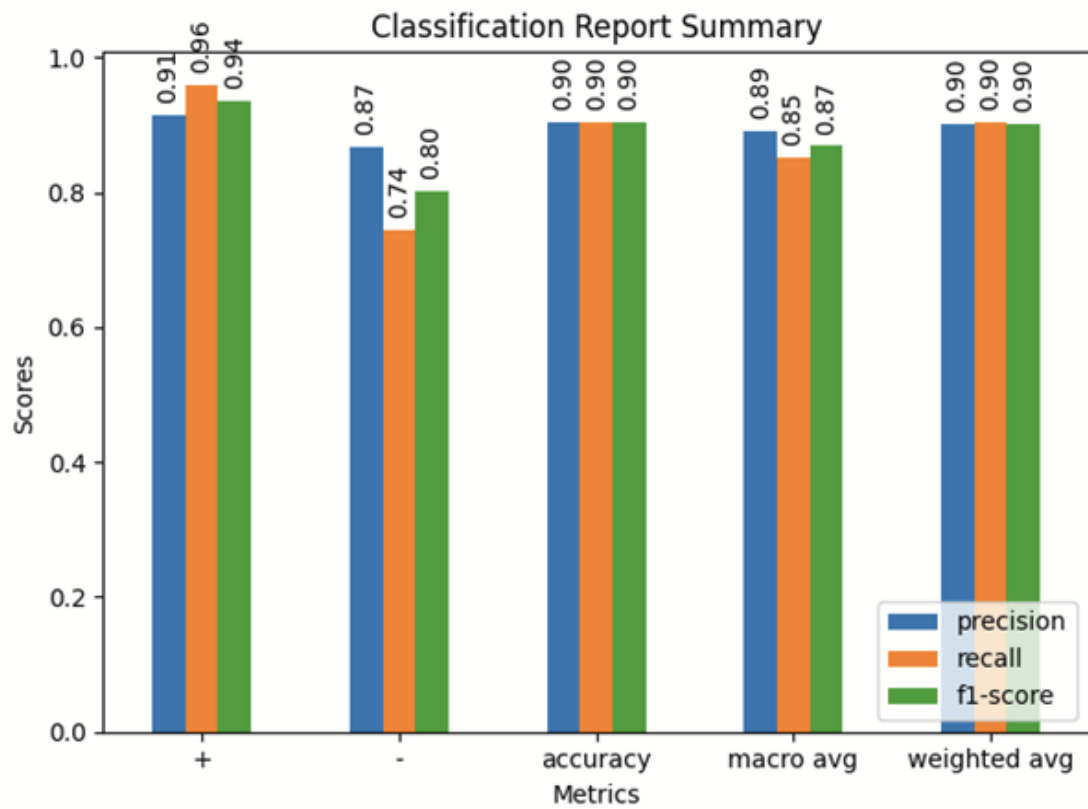


FIGURE 5.9 – Rapport Résultat du model SVM

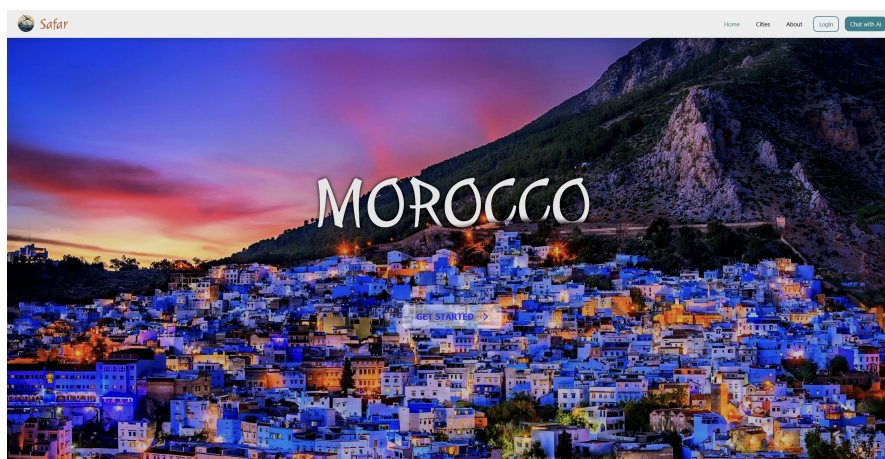


FIGURE 5.10 – Capture d’écran de la page d’accueil

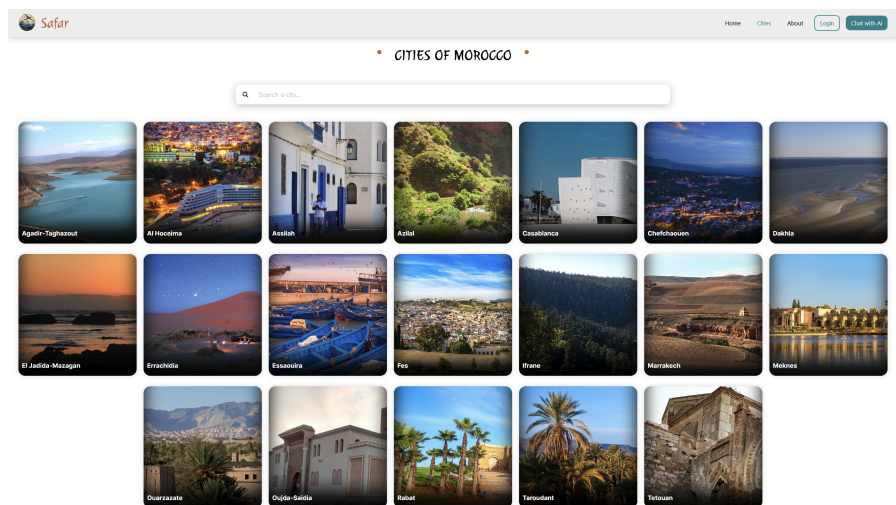


FIGURE 5.11 – Capture d'écran de la page cities

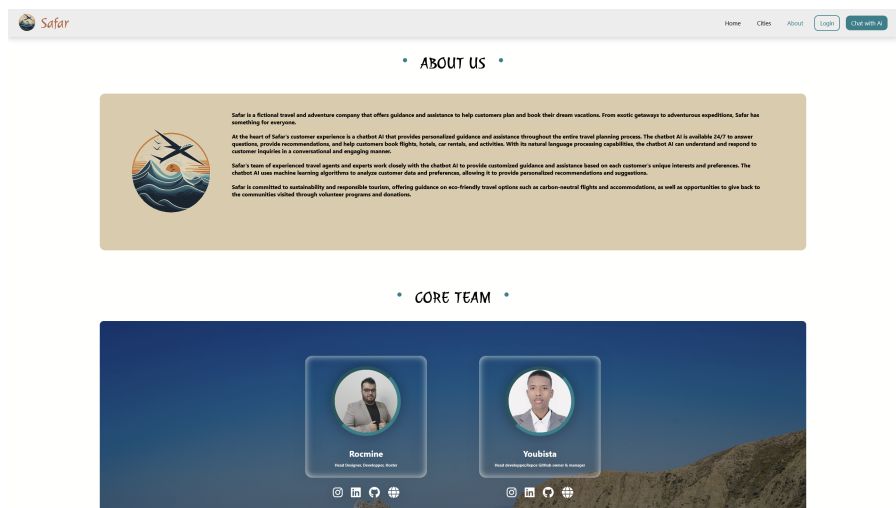


FIGURE 5.12 – Capture d'écran de la page about

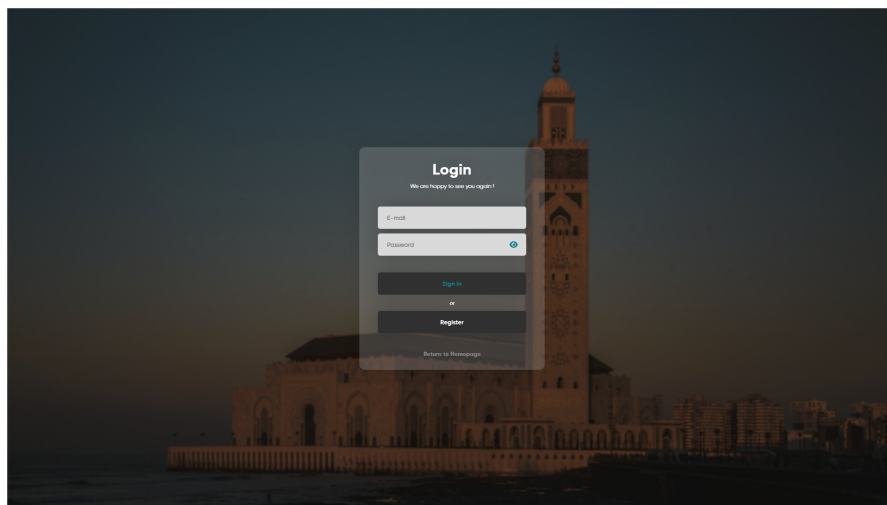


FIGURE 5.13 – Capture d'écran de la page login

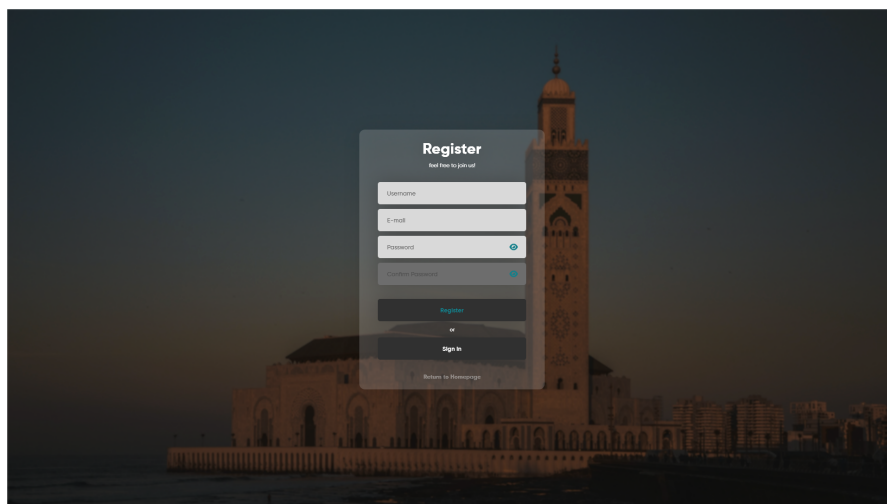


FIGURE 5.14 – Capture d'écran de la page register

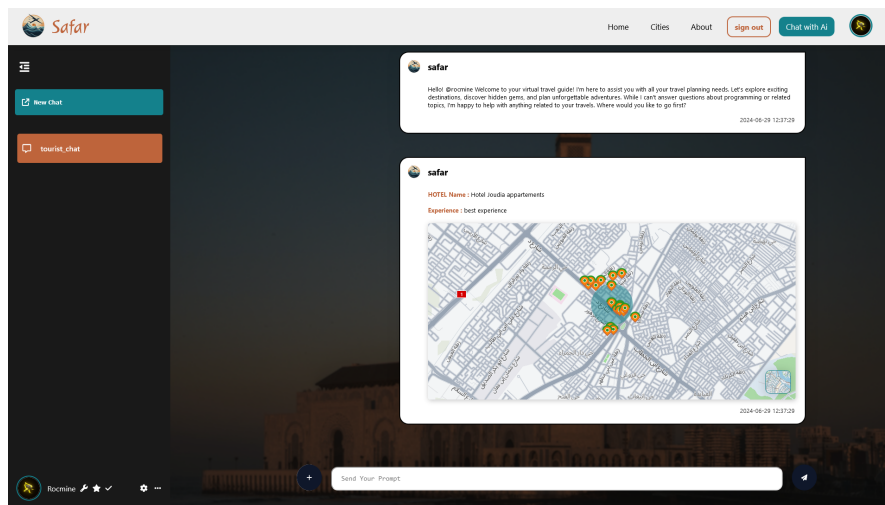


FIGURE 5.15 – Capture d'écran de la page chatbot

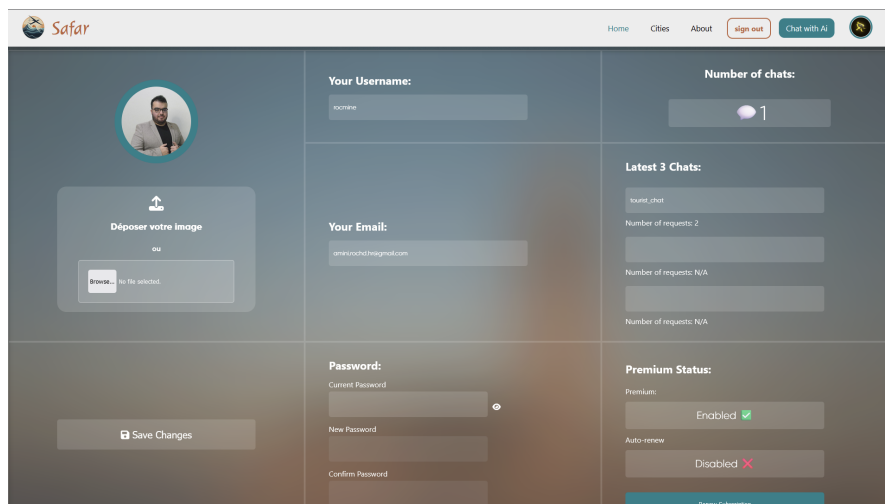


FIGURE 5.16 – Capture d'écran de la page profile

Webographie

1. Site officiel de Python. <https://www.python.org>
2. Documentation de TensorFlow. <https://www.tensorflow.org>
3. Site officiel de VS Code. <https://code.visualstudio.com>
4. RocmineLabs Servers. <https://rocmine.net>
5. Article sur l'analyse des sentiments avec BERT. Miller, A., et al. "BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding." *arXiv preprint arXiv :1810.04805*, 2018.