

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire  
وزارة التعليم العالي والبحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



N°Réf:.....

Centre Universitaire  
Abd Elhafid Boussouf Mila

Institut des Sciences et Technologie

Département de Mathématiques et Informatique

## Mémoire préparé en vue de l'obtention du diplôme de Master

En: Informatique

Spécialité: Sciences et Technologies de l'Information et de la  
Communication (STIC)

### *Thème*

**Analyse Sémantique des Opinions dans Le  
Domaine Économique.**

Préparé par:

**BALI Dina Sirine**

Soutenue devant le jury:

Président

**Merabet Adil**

**Grade MAA**

Examineur

**Boumssata Meriem**

**Grade MAA**

Encadré par

**HADJI Atmane**

**Grade MAA**

Année Universitaire:2023/2024

# **Remerciement**

*Je remercie, Tout d'abord, ALLAH pour la volonté, la force, la santé et la patience qu'il m'a donné afin de réaliser ce travail.*

*Je tiens à adresser mes plus chaleureux remerciements à mon promoteur Mr HADJI .A, pour sa présence à tout moment, sa confiance et sa patience, ainsi que pour ses remarques, ses conseils précieux et sa disponibilité, tout au long de la réalisation de ce travail.*

*Je remercie également les membres du jury de m'avoir honorée avec leurs présences et pour avoir jugé mon travail.*

*J'exprime ma gratitude à l'ensemble du corps enseignant, administratif du département informatique, pour leur disponibilité et leur gentillesse.*

*J'exprime mon infinie gratitude à mes chers parents en reconnaissance de leurs sacrifices, soutien et encouragements.*

*Finalement, nos sincères remerciements à tous ceux qui ont contribué de près ou de loin à la réalisation de ce modeste travail.*

*Dina Sirine*

# Dédicace

*Avec un énorme plaisir, un cœur ouvert et une immense joie, que je  
dédie ce modeste travail:*

*À mon très cher papa « **Cherif** », qui est ma raison de mon existence et  
ma source de puissance*

*À ma chère maman « **Samia** », la source de tendresse, qu'Allah leur  
fasse miséricorde*

*À mon deuxième père et à la personne la plus proche de mon cœur,  
mon cher frère « **Monsaf** »*

*À mon cher frère affectueux « **Imad** »*

*À ma deuxième maman qui m'a toujours soutenue « **Karima** »*

*À mes amies les plus chères et les plus précieuses « **Balkis, Asma** »  
qu'Allah les protèges*

*Et à tous mes autres chères amies sans exception*

*À mon encadreur **Hadji .A** pour tous ses conseils et son aide jusqu'à la  
dernière minute, que Dieu le garde en bonne Santé*

*À toute ma chère famille « **Bali** »*

*Je veux remercier tous ceux qui m'ont soutenue tout au long de ma  
maladie et qui m'ont aidée à continuer*

**Dina Sirine**

## Résumé

Les réseaux sociaux constituent une source précieuse d'information et d'extraction d'opinions, où la majorité des internautes partagent leurs sentiments et avis. L'analyse sémantique des opinions est un outil essentiel pour extraire des informations précieuses à partir de grandes quantités de textes. Elle combine des techniques linguistiques et des méthodes d'apprentissage automatique pour fournir une compréhension approfondie des sentiments et des opinions des utilisateurs. Cette méthode permet de déterminer si les opinions ou sentiments exprimés dans un texte sont positifs, négatifs ou neutres. Elle est utilisée dans divers domaines, notamment le marketing, la gestion de la réputation, les études de marché, l'analyse politique, la finance et l'économie. Dans le domaine économique, cette technique permet d'analyser les sentiments exprimés dans les actualités économiques, les réseaux sociaux, les rapports financiers et autres sources textuelles pertinentes pour comprendre l'état et les tendances de l'économie.

L'objectif de cette étude est de présenter des méthodes visant à améliorer la détection sémantique des opinions en utilisant l'ontologie et la machine learning. L'ontologie joue un rôle essentiel dans l'analyse sémantique des opinions en fournissant un cadre formel pour représenter la connaissance du domaine. L'utilisation d'ontologies permet d'obtenir une analyse plus précise. En intégrant des techniques de machine learning, telles que l'algorithme de l'arbre de décision, nous pouvons améliorer la qualité des résultats obtenus, offrant ainsi un système plus robuste et précis pour l'extraction des opinions.

Dans cette étude, nous avons présenté notre approche d'un système de détection d'opinions sur les réseaux sociaux (Facebook) concernant la monnaie numérique Bitcoin. Nous avons détaillé les différents résultats obtenus en utilisant deux méthodes : la première basée sur les ontologies et la seconde sur la machine learning.

**Mots-clés** : Analyse sémantique des opinions, réseaux sociaux, ontologie, Bitcoin, machine learning.

## Abstract

Social networks constitute a valuable source of information and opinion extraction, where the majority of internet users share their feelings and opinions. Sentiment analysis is an essential tool for extracting valuable information from large amounts of text. It combines linguistic techniques and machine learning methods to provide a deep understanding of users' sentiments and opinions. This method allows for determining whether the opinions or sentiments expressed in a text are positive, negative, or neutral. It is used in various fields, including marketing, reputation management, market research, political analysis, finance, and economics. In the economic field, this technique is used to analyze sentiments expressed in economic news, social networks, financial reports, and other relevant textual sources to understand the state and trends of the economy.

The objective of this study is to present methods aimed at improving the semantic detection of opinions using ontology and machine learning. Ontology plays a crucial role in sentiment analysis by providing a formal framework for representing domain knowledge. Using ontologies in sentiment analysis allows for more accurate analysis. By integrating machine learning techniques, such as the decision tree algorithm, we can enhance the quality of the results obtained, offering a more robust and precise system for opinion extraction.

In this study, we present our approach to an opinion detection system on social networks (Facebook) regarding the digital currency Bitcoin. We detail the various results obtained using two methods: the first based on ontologies and the second on machine learning.

**Keywords:** Sentiment analysis, social networks, ontology, Bitcoin, machine learning.

تشكل وسائل التواصل الاجتماعي مصدرًا قيمًا للمعلومات واستخلاص الآراء، حيث يتشارك غالبية مستخدمي الإنترنت مشاعرهم وآراءهم. يعتبر التحليل الدلالي للآراء أداة أساسية لاستخراج معلومات قيمة من كميات كبيرة من النصوص. يجمع بين التقنيات اللغوية وطرق التعلم الآلي لتوفير الفهم العميق والدقيق لمشاعر وآراء المستخدمين. تسمح هذه الطريقة بتحديد ما إذا كانت الآراء أو المشاعر المعبر عنها في النص إيجابية أو سلبية أو محايدة. تُستخدم هذه الطريقة في مجالات متنوعة، بما في ذلك التسويق وإدارة السمعة وأبحاث السوق والتحليل السياسي والمالية والاقتصاد. في المجال الاقتصادي، تُستخدم هذه التقنية لتحليل المشاعر المعبر عنها في الأخبار الاقتصادية ووسائل التواصل الاجتماعي والتقارير المالية وغيرها من المصادر النصية ذات الصلة لفهم حالة الاقتصاد واتجاهاته.

يهدف هذا البحث إلى عرض طرق تهدف إلى تحسين الكشف الدلالي عن الآراء باستخدام الأنطولوجيا والتعلم الآلي. تلعب الأنطولوجيا دورًا حيويًا في التحليل الدلالي للآراء من خلال توفير إطار رسمي لتمثيل المعرفة في المجال. يتيح استخدام الأنطولوجيات في التحليل الدلالي للآراء الحصول على تحليل أكثر دقة. من خلال دمج تقنيات التعلم الآلي، مثل خوارزمية شجرة القرار، يمكننا تحسين جودة النتائج المحصلة، مما يوفر نظامًا أكثر قوة ودقة لاستخلاص الآراء.

في هذه الدراسة، قدمنا نهجنا لنظام الكشف عن الآراء على وسائل التواصل الاجتماعي (فيسبوك) بشأن العملة الرقمية بيتكوين. لقد فصلنا النتائج المختلفة المحصلة باستخدام طريقتين: الأولى تعتمد على الأنطولوجيا والثانية على التعلم الآلي.

**الكلمات المفتاحية:** التحليل الدلالي للآراء، وسائل التواصل الاجتماعي، الأنطولوجيا، بيتكوين، التعلم الآلي.

# Table de Matière

Remerciement .....	I
Dédicace.....	II
Résumé.....	III
Abstract.....	IV
ملخص .....	V
Table de matière .....	VI
Liste des figures .....	IX
Liste des tableaux .....	X
Liste des sigles et acronymes .....	XI
Introduction générale.....	1

## Chapitre 01 : Analyse sémantique des opinions

1.1 Introduction.....	3
1.2 Analyse des sentiments.....	3
1.2.1 Définition.....	3
1.2.2 Sentiments.....	3
1.2.3 Opinions.....	4
1.2.3.1 Type d'opinion.....	4
1.2.4 Traitement du langage naturel.....	4
1.3 Domaines d'applications de l'analyse des sentiments.....	4
1.3.1 Application en sociologie, psychologie et analyse des sentiments politiques.....	5
1.3.2 Application en marketing.....	5
1.3.3 Application en soins de santé.....	5
1.3.4 Applications en économie et finance.....	5
1.4 Niveaux d'analyse des sentiments.....	7
1.5 Approches d'analyses des sentiments.....	8
1.5.1 Approche lexicale.....	8
1.5.2 Approche d'apprentissage automatique.....	9
1.5.3 Approche hybride.....	10
1.5.3.1 Approche basée dictionnaire.....	10
1.5.3.2 Approche basé sur corpus .....	11
1.6 Difficultés de l'analyse des opinions.....	12
1.7 Conclusion.....	12

## Chapitre 02 : Les ontologies et l'analyse sémantique des opinions.

2.1 Introduction.....	13
2.2 Ontologies.....	13
2.3 Classifications des ontologies.....	14
2.3.1. Selon l'objet de conceptualisation.....	15
2.3.1.1. Ontologie pour la représentation de connaissances.....	15
2.3.1.2. Ontologies supérieures (Upper or Top Level Ontologie).....	15
2.3.1.3. Ontologie Générique (Generic ontology).....	15
2.3.1.4. Ontologies de domaine.....	16
2.1.3.5. Ontologies de tâche.....	16

2.1.3.6. Ontologies d'application.....	16
2.3.2. Selon le degré de formalisme de la représentation.....	16
2.3.2.1 Hautement informelle.....	16
2.3.2.2 Semi-informelle.....	16
2.3.2.3 Semi-formelle.....	16
2.3.2.4 Rigoureusement formelle.....	16
2.3.3. Selon le niveau de détail.....	16
2.3.3.1 Granularité fine.....	17
2.3.3.2 Granularité large.....	17
2.3.3 Selon le niveau de complétude.....	17
2.3.4.2 Niveau Sémantique.....	17
2.3.4.2 Niveau Référentiel.....	17
2.3.4.3 Niveau Opérationnel.....	17
2.4 Avantages de l'ontologie.....	17
2.5 Utilisation de l'ontologie.....	18
2.6 Conclusion.....	18

### **Chapitre 03 : L'analyse des opinions dans le domaine économique**

3.1 Introduction.....	19
3.2 Bitcoin.....	19
3.2.1. Définition.....	19
3.3 Fonctionnement de Bitcoin.....	20
3.3.1 Blockchain.....	20
3.3.2 Processus de transaction.....	20
3.3.3 Nœuds.....	20
3.3.4 Proof-of-work (PoW).....	20
3.3.5 Minage.....	21
3.4 Caractéristiques de Bitcoin.....	22
3.4.1 Décentralisation.....	22
3.4.2 Quantité limitée.....	22
3.4.3 Transparence.....	22
3.4.4 Anonymat relatif.....	22
3.4.5 Facilité de mise en place.....	22
3.5 Avantages de Bitcoin.....	23
3.6 Inconvénients de Bitcoin.....	23
3.7 Travaux antérieurs .....	24
3.7.1 The Relationship Between Social Media Sentiment and Bitcoin Price Volatility .....	24
3.7.2 Sentiment-Driven Cryptocurrency Price Prediction... ..	25
3.8 Conclusion.....	27

### **Chapitre 04 : Conception d'un système de détection sémantique des opinions**

4.1. Introduction.....	28
4.2. Notre Approche Proposée.....	28
4.3. Architecture de modèle proposé.....	29
4.4 Étapes de l'approche proposée... ..	31
4.4.1 Étape de collection des données.....	31
4.4.2 Prétraitement.....	31
4.4.3 Étape de création de l'ontologie.....	31

4.4.4	Étape d'analyse d'un texte.....	33
4.4.4.1	Tokeniser.....	33
4.4.4.2	Sentence Splitter.....	34
4.4.4.3	Part Of Speech Tagger.....	34
4.4.4.4	Analyseur Morphologique.....	34
4.4.4.5	Onto Root Gazetteer.....	35
4.4.5	Machine Learning .....	36
4.4.6	Conception d'un modèle basé sur machine Learning.....	36
4.4.6.1	Notre proposition .....	36
4.4.6.2	Algorithme de l'arbre de décision.....	38
4.4.6.3	Arbre de Décision Tree Classifier.....	38
4.5	Conclusion.....	39

## **Chapitre 05 : Implémentation et Evaluation**

5.1	Introduction.....	40
5.2	Protégé.....	40
5.2.1	Création de l'ontologie.....	41
5.3	GATE.....	43
5.4	ANNIE.....	44
5.4.1	Tokeniser (TokeniserPR).....	45
5.4.2	Gazetteer (GazetteerPR).....	45
5.4.3	Sentence Splitter.....	45
5.4.4	POS Tagger.....	45
5.4.5	Named-Entity transducer (NE transducer).....	45
5.5	Différentes méthodes appliquées.....	45
5.5.1	Méthode par défaut.....	45
5.5.2	Méthode basé sur l'ontologie.....	46
5.5.3	Méthode basé sur la machine Learning.....	47
5.6	Evaluation.....	48
5.7	Conclusion.....	51
	Conclusion générale.....	53

## Liste Des Figures

<b>Figure 1.1 :</b> Domaines d'applications de l'analyse des sentiments.....	6
<b>Figure 1.2:</b> Les niveaux de l'analyse des sentiments.....	7
<b>Figure 1.3:</b> Les approches d'analyses des sentiments.....	8
<b>Figure 2.1:</b> Exemple d'une simple ontologie.....	14
<b>Figure 2.2 :</b> Typologies d'ontologies selon quatre dimensions de classification.....	15
<b>Figure 3.1 :</b> un exemple de fonctionnement de Bitcoin.....	21
<b>Figure 3.2 :</b> Le schéma proposé de cette étude.....	25
<b>Figure 3.3 :</b> la méthodologie utilisée dans cette recherche.....	26
<b>Figure 4.1:</b> Architecture générale du système proposé.....	30
<b>Figure 4.1:</b> Exemple d'un Tokeniser.....	33
<b>Figure 4.2:</b> Exemple d'un Sentence Splitter.....	34
<b>Figure 4.3:</b> Exemple d'un Part Of Speech Tagger.....	35
<b>Figure 4.4:</b> Building Ontology Resource Root (OntoRoot) Gazetteer from the Ontology .....	36
<b>Figure 4.5:</b> Schéma de fonctionnement de la machine Learning.....	37
<b>Figure 4.6:</b> un exemple de l'arbre de décision.....	38
<b>Figure 5.1:</b> Protégé Capture d'écran .....	40
<b>Figure 5.2 :</b> Classes d'ontologie de haut niveau.....	42
<b>Figure 5.3 :</b> Hiérarchie de propriété d'objet.....	42
<b>Figure 5.4 :</b> Exemple de restrictions de rôle.....	43
<b>Figure 5.5 :</b> Liste de certaines instances d'ontologie.....	43
<b>Figure 5.6:</b> Exemple de l'interface dans GATE.....	44
<b>Figure 5.7 :</b> Exécution Opinion Bitcoin par Default.....	46
<b>Figure 5.8 :</b> Exécution d'ontologie (lookup) (Opinion Bitcoin).....	46
<b>Figure 5.9 :</b> Les packages de Python.....	47
<b>Figure 5.10:</b> L'exécution de la machine Learning.....	48

## Liste des Tableaux

<b>Tableau 4.1:</b> Classes, Sub-classes et sous sous classes de l'Ontologie .....	32
<b>Tableau 5.1 :</b> Evaluation Opinion Positive .....	50
<b>Tableau 5.2 :</b> Evaluation Opinion Négative .....	50
<b>Tableau 5.3 :</b> Evaluation Opinion Neutre.....	51

## Liste des Sigles et Acronymes

<b>ANNIE</b>	A Nearly New Information Extraction
<b>NLP</b>	Natural Language Processing
<b>SVM</b>	Support Vector Machine
<b>TALN</b>	Traitement Automatique de Langues Naturel
<b>OWL</b>	Ontology Web Language
<b>GATE</b>	General Architecture for Text Engineering
<b>ML</b>	Machine Learning
<b>TAO</b>	Transitioning Applications to Ontologies
<b>NL</b>	Langue Nature
<b>EN</b>	Entité nommée
<b>PoW</b>	Proof-of-Work
<b>TF-IDF</b>	Term Frequency-Inverse Document Frequency

### Introduction générale

Au cours des dernières années, les réseaux sociaux sont devenus populaires grâce au fait qu'ils ont réussi à s'imposer parmi les services les plus prisés sur Internet. Le Web offre désormais une vaste source d'informations où chacun peut exprimer ses opinions et découvrir celles des autres. Cette croissance spectaculaire des réseaux sociaux a conduit à l'émergence de l'analyse des sentiments, il s'agit de l'étude informatique des opinions, sentiments et émotions exprimés dans le texte, elle trouve ainsi des applications dans différents domaines tels que le marketing, la finance, l'économie et la politique entre autres.

L'analyse sémantique des opinions, ou analyse des sentiments, est un domaine de recherche en pleine expansion qui vise à extraire et analyser les émotions, les opinions et les attitudes exprimées. Dans le domaine économique, cette approche permet de comprendre les perceptions des consommateurs, des investisseurs et du public en général, à l'égard de différents sujets économiques, des produits aux politiques économiques en passant par les fluctuations du marché. L'analyse des sentiments s'appuie sur des techniques de traitement du langage naturel (NLP) pour convertir des données textuelles brutes en informations utilisables.

Dans le domaine économique en particulier l'analyse des opinions, concernant le Bitcoin, la célèbre crypto-monnaie, offrant une abondance de données pour l'analyse. En étudiant les sentiments et les opinions associés à ce dernier, il est possible de détecter des tendances, d'anticiper les mouvements de marché et de mieux comprendre les facteurs psychologiques influençant les investisseurs et les utilisateurs. Cette capacité à analyser et interpréter les opinions publiques sur Bitcoin illustre l'utilité et la puissance de l'analyse sémantique des opinions dans le domaine économique.

L'analyse sémantique des opinions est un processus essentiel et incontournable lorsque les opinions sont converties en fichiers électroniques, que ce soit sur des sites web, des médias ou des réseaux sociaux. Cependant, l'extraction de ces opinions sans utiliser d'outils supplémentaires ni préciser exactement le contenu recherché par l'utilisateur peut conduire à des résultats insatisfaisants ou inefficaces. En effet, cela peut parfois entraîner des similitudes non souhaitées ou des erreurs complètes.

Pour résoudre cette problématique, le système d'extraction fondé sur l'ontologie se distingue des structures d'indexation de données statiques traditionnelles. En effet, il permet de reconnaître le type d'opinions extraites en les associant à leurs descriptions sémantiques dans les ontologies. De plus, l'intégration de la machine Learning peut améliorer la précision de cette extraction en utilisant des modèles entraînés à reconnaître les contextes dans lesquels les opinions sont exprimées, ce qui conduit à des résultats plus fiables et pertinents.

Dans cette proposition, l'objectif de ce travail est d'améliorer les techniques de la recherche et de l'extraction d'opinions. Notre approche principale consiste à identifier les opinions contenues dans un ensemble de documents ou de textes disponibles en ligne.

Pour développer cet aspect nous avons subdivisé notre travail en cinq chapitres :

- Le premier chapitre aborde l'analyse sémantique des opinions en présentant une définition générale et les principes de base. Nous explorerons ensuite les différents domaines de l'analyse des sentiments, ainsi que les niveaux de l'analyse des sentiments, puis nous allons citer quelque difficulté de l'analyse des opinions. Enfin, nous terminons ce chapitre par présenter les approches de l'analyse des sentiments.
- Dans le deuxième chapitre, nous explorerons les ontologies, leurs classifications, puis nous allons donner les avantages et l'utilisation.
- Dans le troisième chapitre, nous allons présenter la définition de la monnaie numérique Bitcoin, ainsi que son fonctionnement et ses caractéristiques, puis nous donnerons ses avantages et inconvénients. Nous terminons ce chapitre par les travaux déjà réalisés.
- Dans le quatrième chapitre, nous introduisons un modèle de conception en exposant et décrivant notre méthode ainsi que l'architecture détaillée et les différentes phases.
- Dans le cinquième chapitre, nous allons définir l'outil Protégé pour la création de l'ontologie, le GATE développer et ses composants ANNIE et la machine Learning pour les comparer avec l'ontologie. Nous allons expliquer par la suite deux méthodes appliqués, la méthode d'ontologie est appliquée sur GATE et la méthode de la machine Learning est appliqué en utilisant l'algorithme de l'arbre de décision sur Python. Nous allons aussi donner les paramètres de Python et les avantages de l'algorithme utilisé, ainsi que l'évaluation de chaque méthode et finalement nous allons faire une comparaison entre les deux méthodes.

**Chapitre 1 :**  
Analyse sémantique des  
Opinions

## 1.1 Introduction

L'analyse sémantique des opinions, aussi connue sous le nom d'analyse des sentiments, a gagné en popularité ces dernières années. Initialement, cette technique était principalement employée par les entreprises pour des fins commerciales, permettant de mieux comprendre les perceptions des consommateurs et d'adapter leurs stratégies marketing en conséquence. L'analyse des sentiments représente un véritable atout pour la surveillance sociale, offrant une alternative fiable aux méthodes d'enquête traditionnelles. Elle permet de recueillir et d'analyser en temps réel les opinions des utilisateurs exprimées sur diverses plateformes en ligne. De nos jours, cette méthode n'est plus uniquement réservée aux entreprises. Elle est également utilisée pour mesurer l'opinion publique sur des sujets d'actualité, notamment via les réseaux sociaux comme Twitter ou les blogs. Grâce à cette analyse, il devient possible de saisir les tendances, les émotions et les réactions des internautes, fournissant ainsi des insights précieux pour les décideurs et les chercheurs.

Dans ce chapitre, nous allons présenter les concepts et définitions fondamentaux liés à l'analyse sémantique des opinions, explorer les domaines d'application variés et décrire les différentes approches méthodologiques utilisées dans ce domaine.

## 1.2 Analyse des sentiments

### 1.2.1 Définition

L'analyse des sentiments consiste à examiner les émotions et les opinions exprimées par les individus sur internet, ainsi que les avis fournis par les clients aux entreprises. En termes plus généraux, cette discipline, également appelée exploration d'opinions, fait appel à des techniques d'exploration de données et de traitement du langage naturel (NLP) pour identifier, extraire et synthétiser des informations et des opinions à partir d'un vaste corpus de textes disponibles sur le Web. Grâce à l'analyse des sentiments, il est possible de suivre les perceptions et ressentis exprimés en ligne, que ce soit dans des articles de blog, des commentaires, des critiques ou des tweets, afin de déterminer si ces opinions sont perçues de manière positive, négative ou neutre [1] .

### 1.2.2 Sentiments

Les sentiments sont comme les jugements que l'on porte sur un objet ou un sujet, ce jugement est caractérisé par la polarité et l'intensité. Pour que la polarité soit des valeurs positives ou négatives ou un mélange des deux, tandis que l'intensité montre le degré de positivité ou de négativité, allant de faible à fort [2].

### 1.2.3 Opinions

L'opinion désigne un avis que l'on exprime à l'égard d'une personne, d'un être vivant, d'un phénomène, d'un fait, d'un objet ou d'une chose. On peut la juger positive ou négative, cela varie en fonction de la nature de l'individu en fonction de son caractère, de ses émotions et de son comportement [3].

#### 1.2.3.1 Type d'opinion

Les opinions pouvant être positives, négatives ou neutres, il existe deux catégories principales d'opinion :

- **Opinion usuelle ou comparative**

- ✓ **Opinion usuelle** : est un simple avis qui peut être visé d'une manière directe ou indirecte sur un sujet principal. Les opinions directes sont les plus exploitées dans les études.
- ✓ **Opinion comparative** : est un sous-domaine de l'exploration d'opinion qui traite de l'identification et de l'extraction d'informations exprimées sous forme comparative.

- **Opinion explicite ou implicite**

- ✓ **Opinion explicite** : est un extrait sans aucun mot d'opinion est d'un avis subjectif.
- ✓ **Opinion implicite** : la clause ne contient aucune opinion, ces avis sont généralement objectifs, c'est la catégorie la moins explorée dans les études [4].

### 1.2.4 Traitement du langage naturel

Le traitement du langage naturel (Natural Language Processing ou NLP en anglais) est défini comme un domaine spécialisé de l'informatique et de l'ingénierie, et fait partie intégrante de l'intelligence artificielle, ancrée dans la linguistique informatique. Le NLP est utilisé pour la création d'applications et de systèmes permettant l'interaction entre les machines et le langage naturel créé par les humains. Cette technologie permet aux ordinateurs de traiter et de comprendre le langage humain de manière naturelle, afin de fournir des résultats utiles [5].

## 1.3 Domaines d'applications de l'analyse des sentiments

De nos jours, l'analyse des sentiments a gagné encore plus de valeur et leur grande diffusion et leur rôle dans la société moderne, représentent l'une des nouveautés les plus intéressantes de ces dernières années, captant l'intérêt des chercheurs, des journalistes, des

entreprises et des gouvernements.

On peut dire que l'analyse des sentiments est régulièrement convoquée dans tout processus décisionnel, tel que :

### **1.3.1 Application en sociologie, psychologie et analyse des sentiments politiques**

Il permet à un individu d'exprimer une opinion (critique) sur quelque chose à l'échelle mondiale, par exemple critiques de films, opinions politiques, une identification de connard dans les commentaires, les actualités, etc. Fait référence à l'émotion derrière une mention sur les réseaux sociaux, c'est un sens d'évaluer le ton de la conversation : la personne est-elle heureuse ou en colère. En politique, nous pouvons analyser les tendances, identifier les préjugés idéologiques, cibler les publicités/messages en conséquence et évaluer les opinions du public/clients électeur. En sociologie, la diffusion des idées à travers des groupes est un concept. La psychologie et l'analyse des sentiments fournissent une plate-forme d'amélioration enquêtes/expériences psychologiques à partir de textes NL (langue Nature)[6].

### **1.3.2 Application en marketing**

Les réseaux sociaux sont aujourd'hui devenus une plateforme unique pour interagir avec les clients. En utilisant l'analyse de sentiment, il est facile d'aborder le marketing à un niveau tout autre, où les entreprises ont réalisé que les émotions des réseaux sociaux influencent l'image de leur marque. À l'inverse, les outils d'analyse des sentiments permettent aux experts en marketing de mesurer leur efficacité et de soutenir les consommateurs qui cherchent un produit ou un service [7].

Le domaine de l'Opinion Manning est devenu un problème majeur pour toute entreprise désireuse de mieux comprendre les facteurs qui contribuent à la satisfaction et à l'insatisfaction des clients [8].

### **1.3.3 Application en soins de santé**

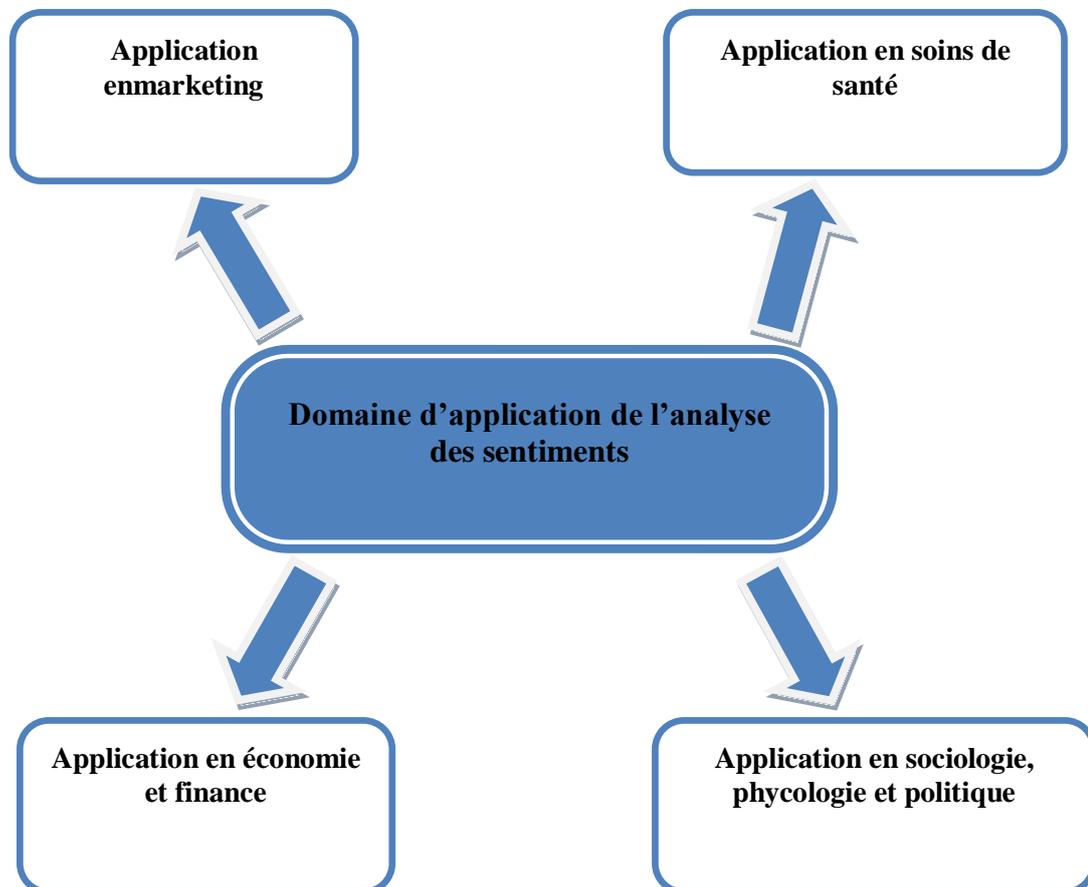
Les blogs médicaux sont omniprésents sur internet de nos jours. Ils contiennent des informations sur divers problèmes médicaux et de santé tels que les maladies, les traitements médicaux et pharmaceutiques. En raison de l'expérience liée à la santé et ces pages fournissant des antécédents médicaux aux médecins et aux patients, des outils d'analyse des sentiments doivent être développés pour le domaine médical [7].

### **1.3.4 Applications en économie et finance**

L'analyse des sentiments peut également être utilisée en finance. Afin que les investisseurs puissent facilement suivre leurs activités, collecter et surveiller leurs données de

sentiment en temps réel. L'analyse des émotions aide les investisseurs à obtenir plus facilement des informations commerciales [7]. Les marchés financiers accordent une grande importance à l'analyse des sentiments car elle permet d'obtenir des informations précieuses sur la perception des investisseurs et les fluctuations du marché. Grâce à l'analyse des émotions, les investisseurs et les traders peuvent prendre des décisions éclairées en analysant les sentiments et les opinions exprimés dans les articles de presse, les publications sur les réseaux sociaux et d'autres sources de qui permet de saisir le sentiment global du marché [9].

L'analyse des sentiments occupe une place essentielle dans le domaine économique en offrant des renseignements précieux sur les opinions, les émotions et les tendances des consommateurs, des investisseurs et des acteurs du marché (Figure 1.1).



**Figure 1.1** : Domaines d'applications de l'analyse des sentiments.

## 1.4 Niveaux d'analyse des sentiments

Il existe trois niveaux d'analyse : le niveau du document (Message level ou Document level), le niveau de la phrase (Sentence level) et le niveau des aspects (Entity and aspect level) (figure 1.2).

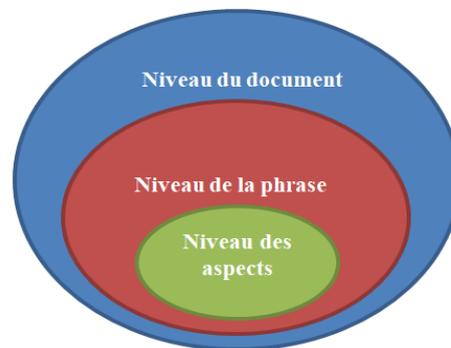


Figure 1.2 : Niveaux de l'analyse des sentiments.

### ▪ Niveau du document

À ce stade, la mission consiste à évaluer si un document d'opinion exprime un sentiment positif ou négatif. Par exemple, lorsqu'un avis sur un produit est donné, le système évalue si l'avis exprime une opinion globale positive ou négative sur le produit. On appelle généralement cette tâche, la classification des émotions au sein du document [10]. La classification la plus simple consiste à considérer l'ensemble des documents de texte d'opinion comme une unité d'information élémentaire. Le document est supposé exprimer une opinion sur un seul sujet. Cette méthode n'est pas appropriée lorsque le document renferme des opinions sur divers sujets tels que dans les forums et les blogs [11].

### ▪ Niveau de la phrase

À ce stade, la mission consiste à évaluer si chaque phrase a exprimé une opinion positive, négative ou neutre. En général, être neutre ne signifie pas avoir d'opinion. La classification de la subjectivité est étroitement liée à l'analyse, qui différencie les phrases qui expriment des informations, des phrases qui expriment des opinions et des vues subjectives [3]. L'examen des émotions dans la phrase est l'examen le plus précis du document. Dans cette optique, on évalue la polarité de chaque phrase car chaque phrase est perçue comme une entité distincte et peut avoir une opinion différente.

### ▪ Niveau d'entité et d'aspect

La recherche de documents et de phrases ne révèle pas exactement ce que les individus aiment et n'aiment pas. Une analyse détaillée est effectuée au niveau des aspects. Auparavant, les niveaux d'apparence étaient appelés niveaux fonctionnels. Le niveau aspect ne se

concentre pas sur la forme du langage mais directement sur les idées elles-mêmes. L'importance d'identifier les cibles d'opinion permet également de mieux comprendre la problématique de l'analyse des sentiments [10].

### 1.5 Approches d'analyses des sentiments

Les techniques de classification des sentiments peuvent être grossièrement divisées en une approche d'apprentissage automatique, une approche basée sur le lexique et une approche hybride (figure 1.4).

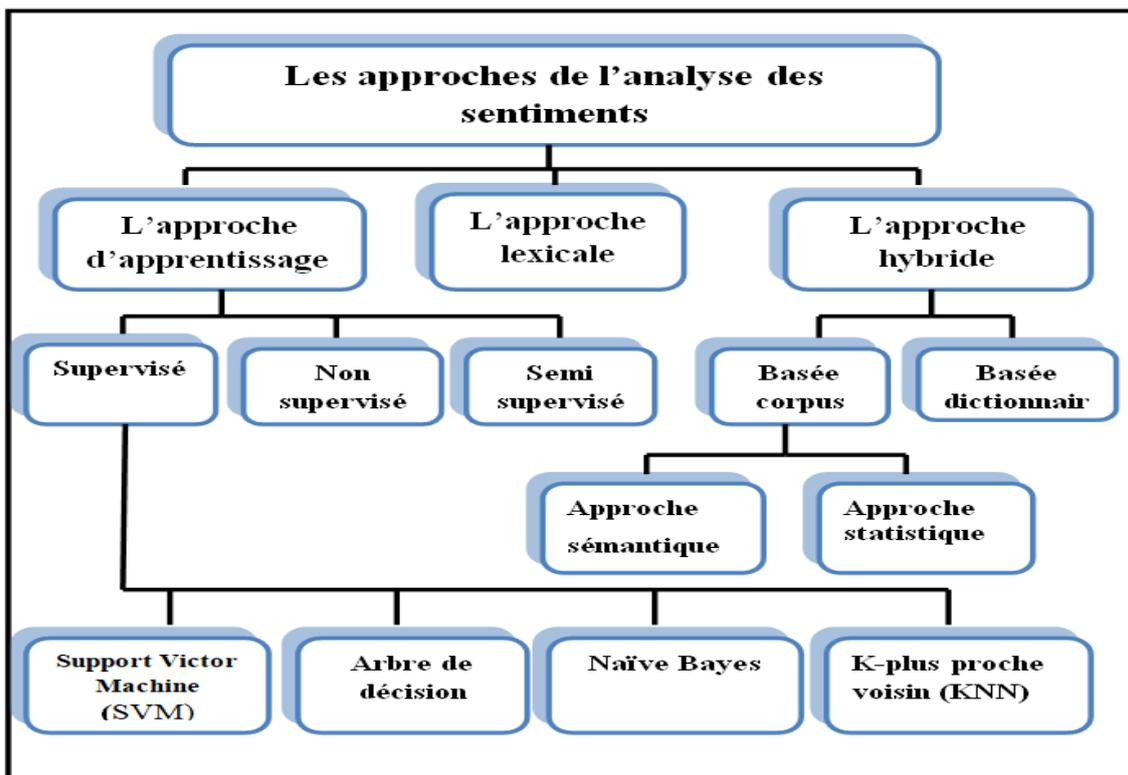


Figure 1.3 : Approches d'Analyses des Sentiments.

#### 1.5.1 Approche lexicale

Cette méthode se base sur des dictionnaires, c'est-à-dire un dictionnaire prédéfini où chaque terme est lié à une polarité : principalement positive, négative et parfois neutre. La majorité des mots sont des verbes (comme "multiplier") et des adjectifs (comme "insuffisant"), ainsi que des noms communs (comme "réduction") et des adverbes (comme "horrible"). La méthode utilisant des mots offre l'avantage de garantir la clarté des critères de classification. Il est relativement simple de voir quelles phrases du texte expriment quelle polarité à partir de quels (s) mots. On peut ajouter, modifier ou supprimer des termes dans les lexiques afin d'améliorer les résultats en cas d'erreurs. Toutefois, la création manuelle de ces

mots requiert une expertise manuelle approfondie et une couverture souvent relativement restreinte. Il est donc essentiel d'être extrêmement exhaustif : il est essentiel de lister les termes afin de les repérer [12].

### 1.5.2 Approche d'apprentissage automatique

Le ML (Machine Learning) est une branche de l'intelligence artificielle qui permet aux systèmes de comprendre grâce à ses algorithmes. L'algorithme est basé sur l'idée d'apprendre à partir de données et de faire des prédictions basées sur ces données. Ainsi, les ordinateurs acquièrent la capacité de résoudre des tâches spécifiques sans avoir à les programmer. Il existe trois catégories d'apprentissage automatique: l'apprentissage supervisé, l'apprentissage non supervisé et l'apprentissage semi-supervisé.

L'apprentissage supervisé permet à un ordinateur de prédire de nouvelles données d'entrée après avoir été entraîné avec des données et des résultats spécifiques. Le deuxième type d'apprentissage est l'apprentissage non supervisé, où l'ordinateur reçoit uniquement des données et doit découvrir par lui-même une structure significative, principalement basée sur le regroupement. Cette méthode d'apprentissage repose sur l'observation plutôt que sur l'imitation. L'apprentissage semi-supervisé, quant à lui, combine des données étiquetées (supervisées) et des données non étiquetées (non supervisées) pour entraîner un modèle [13].

- **Apprentissage supervisé**

De nombreux algorithmes d'apprentissage supervisés peuvent être adaptés à l'analyse des sentiments. Parmi ceux-ci, on trouve le classificateur d'entropie maximale, les machines à vecteurs de support et les arbres de décision, entre autres. Ces algorithmes suivent généralement trois étapes principales : la collecte des données, le prétraitement des données, l'entraînement et le suivi des résultats [14].

Les méthodes d'apprentissage supervisées utilisent divers classificateurs pour analyser les sentiments, notamment les suivants :

- ✓ **Support Vector Machine (SVM) classificateur**
- ✓ **Arbres de décision**
- ✓ **Naïve Bayes classificateur**
- ✓ **K-plus proche voisin (KNN) classificateur**
- **Apprentissage non supervisé**

L'apprentissage non supervisé a fait l'objet de nombreuses recherches au sein de la communauté du Machine Learning. Les algorithmes de clustering, de réduction de dimensionnalité et de densité sont fréquemment employés dans divers secteurs:

- ✓ **Médecine**
- ✓ **Traitement de la parole**
- ✓ **Traitement d'images**
- ✓ **Classification de documents**
- ✓ **Analyse des opinions**
- **Apprentissage semi supervisé**

L'apprentissage semi-supervisé représente une catégorie de techniques d'apprentissage automatique qui combine des données étiquetées et non étiquetées [15]. Il est positionné entre l'apprentissage supervisé, qui utilise uniquement des données étiquetées, et l'apprentissage non supervisé, qui s'appuie exclusivement sur des données non étiquetées, ce type d'approche permet d'améliorer considérablement la qualité des modèles d'apprentissage [16].

### **1.5.3 Approche hybride**

L'approche hybride combine l'apprentissage automatique et les méthodes basées sur le lexique (éventuellement intégrées dans des règles linguistiques). Différents classificateurs de sentiment, qu'ils soient fondés sur un lexique ou sur des techniques d'apprentissage, sont utilisés en cascade. Ainsi, lorsqu'un classificateur échoue, le suivant prend le relais pour classifier le texte restant. Une approche hybride peut utiliser des techniques d'apprentissage automatique pour suggérer de nouveaux termes et créer des dictionnaires plus riches et spécifiques au contexte d'un terme. Ces modèles hybrides donnent souvent d'excellents résultats en termes de précision [17].

Les méthodes hybrides sont très faciles à comprendre : il suffit simplement de combiner les deux approches les plus efficaces, celle basée sur des règles et celle automatique. En général, en combinant ces deux approches, les méthodes peuvent augmenter la précision [17].

#### **1.5.3.1 Approche basée dictionnaire**

La stratégie principale de l'approche par dictionnaire consiste à commencer par rassembler manuellement un petit ensemble de mots d'opinion dont l'orientation est connue. Par la suite, cet ensemble est élargi en utilisant des corpus bien établis tels que WordNet ou des thésaurus pour trouver les synonymes et antonymes de ces mots.

Toutefois, un inconvénient majeur de cette méthode est son incapacité à identifier des mots d'opinion dont l'orientation est spécifique à un domaine ou à un contexte particulier [18].

### 1.5.3.2 Approche basé corpus

L'approche basée sur le corpus permet de surmonter le défi et d'identifier des mots d'opinion en fonction de leur orientation contextuelle. Cette méthode repose sur des modèles syntaxiques ou des algorithmes spécifiques utilisés dans les applications d'analyse des sentiments.

Même si un lexique de sentiments généraux peut être créé en utilisant l'approche basée sur le corpus, à condition de disposer d'un corpus très diversifié et étendu, l'approche basée sur le dictionnaire est généralement plus efficace. En effet, un dictionnaire inclut un nombre de mots beaucoup plus important [19].

Cette approche peut être mise en œuvre en utilisant soit une méthode statistique, soit une méthode sémantique.

- **Approche statistique**

L'approche statistique de l'analyse des sentiments, également connue sous le nom d'approche basée sur l'apprentissage automatique, emploie des algorithmes de machine learning pour détecter et catégoriser les sentiments présents dans un texte.

Cette méthode associe des techniques de traitement du langage naturel (NLP) à des algorithmes de machine learning, formés sur de vastes ensembles de données annotées. Le système est entraîné pour relier les entrées textuelles aux résultats correspondants, tels que les opinions des clients avec leur polarité (positive, négative ou neutre) [20].

- **Approche sémantique**

L'approche sémantique se concentre sur l'examen du sens des mots, des phrases et des expressions pour saisir les émotions et les opinions exprimées. Cette méthode s'appuie sur des techniques telles que la classification, l'analyse lexicale et l'apprentissage automatique pour évaluer les sentiments (positifs, neutres, négatifs) dans les textes.

L'analyse sémantique peut être utilisée dans de nombreuses applications, telles que la reconnaissance de la langue naturelle, la traduction automatique et la recherche d'informations pour optimiser une qualité de produit/service perçue [21].

L'approche sémantique repose aussi sur l'utilisation d'une ontologie, qui est un modèle conceptuel explicitement défini avec une sémantique formelle basée sur la logique. L'ontologie permet de capturer et de représenter les relations sémantiques entre les concepts d'un domaine particulier, facilitant ainsi le partage et la réutilisation des connaissances parmi les utilisateurs et les applications [22].

## 1.6 Difficultés de l'analyse des opinions

La classification des textes d'opinion (positive, négative ou neutre) cherche à analyser les émotions exprimées, mais elle rencontre plusieurs obstacles :

- ✓ **Ambiguïté des mots** : Certains mots peuvent avoir une connotation positive ou négative selon le contexte, rendant leur interprétation difficile.
- ✓ **Structures syntaxiques et sémantiques** : La diversité des structures syntaxiques et sémantiques rend l'expression de l'opinion difficile à évaluer. Par exemple, « l'histoire du film est captivante mais les acteurs étaient médiocres » présente des polarités opposées entre les deux parties de la phrase.
- ✓ **Analyse basée sur des "paquets de mots"** : Deux phrases contenant les mêmes ensembles de mots peuvent exprimer des émotions différentes, posant un défi pour une analyse précise.
- ✓ **Langage des internautes** : Le langage utilisé en ligne pose des problèmes spécifiques, tels que l'absence de ponctuation pour marquer la fin des phrases et l'utilisation de mots comme « Good » ou « super ».
- ✓ **Lexique adéquat** : La difficulté à identifier un lexique adéquat pour l'analyse des textes d'opinion complique la tâche.
- ✓ **Association opinion-requête** : Il est difficile d'associer l'opinion à la requête, car une opinion dans un document ne concerne pas toujours la requête spécifique [23].

## 1.7 Conclusion

Dans ce chapitre, nous avons expliqué en détail les concepts de base de notre domaine d'étude, qui sont la définition, l'analyse des sentiments, les sentiments, les opinions et déterminé leurs types et nous avons également expliqué la signification de traitement automatique du langage naturel. De plus, nous avons présenté les différents domaines d'application de l'analyse des sentiments. En outre, nous avons cité les niveaux de l'analyse des sentiments, différentes approches. Finalement nous avons mentionné les difficultés de l'analyse des sentiments.

## **Chapitre 2 :**

# Ontologies et analyse sémantique des opinions

## 2.1 Introduction

Les ontologies suscitent un grand intérêt au sein de multiples communautés de recherche, notamment en ingénierie des connaissances, en recherche d'information, dans les systèmes d'information coopératifs et dans le commerce électronique. Elles jouent un rôle central dans le Web sémantique. Leur développement vise à faciliter le partage et la mise à jour des connaissances et des raisonnements issus de divers secteurs d'activité et de production.

Dans ce chapitre, nous allons présenter les ontologies et exposer en détail les classifications des ontologies, par la suite nous citerons les avantages de l'analyse des sentiments et leur utilisation.

## 2.2 Ontologies

La définition précise du terme « ontologie » est complexe, car il est utilisé dans des domaines variés tels que la philosophie, la linguistique et l'intelligence artificielle. Initialement, ce mot vient de la philosophie et a été créé au XIXe siècle à partir des racines grecques "onto" (l'existence) et "logos" (l'étude), signifiant ainsi "étude de l'être" [24].

L'ontologie, par définition est la spécification explicite des conceptualisations partagées signifie que les types de concepts utilisés et les contraintes sur leur utilisation sont clairement définis et compris par la machine « formel ». « Explicite » implique que ces concepts et contraintes sont bien détaillés. « Partagé » indique que l'ontologie représente des connaissances consensuelles, acceptées par un groupe et non limitées à des individus. La « conceptualisation » se réfère à un modèle abstrait d'un phénomène du monde, identifiant les concepts associés à ce phénomène [25].

Les ontologies peuvent être considérées comme des modèles déclaratifs qui définissent des domaines, représentant les concepts qui y existent, ainsi que leurs propriétés et les relations entre eux [26]. Elles occupent une place cruciale dans le Web sémantique, s'efforçant de saisir la signification du domaine en utilisant les connaissances brutes représentées, ce qui permet aux machines de comprendre les relations entre les concepts d'un domaine. Cette connaissance est généralement présentée sous la forme d'une base de connaissances, accessible aux applications nécessitant l'utilisation et/ou le partage de la connaissance d'un domaine [25]. Elles constituent une base solide pour la communication entre machines, ainsi qu'entre humains et machines, en définissant le sens des objets par les symboles qui les désignent et les caractérisent, et par une représentation structurée ou formelle de leur rôle dans un domaine [27].

L'objectif est de créer un modèle global et exhaustif qui décrit la signification de l'échange d'informations. Pour faciliter le partage et la réutilisation des connaissances, les ontologies sont employées dans le domaine de l'intelligence artificielle. Elles se composent de concepts, de caractéristiques, de relations et de contraintes entre les concepts qui incluent des classes, des attributs, des propriétés et des relations entre les classes.

La figure 2.1, représente une simple ontologie contenant des classes et ses relations.

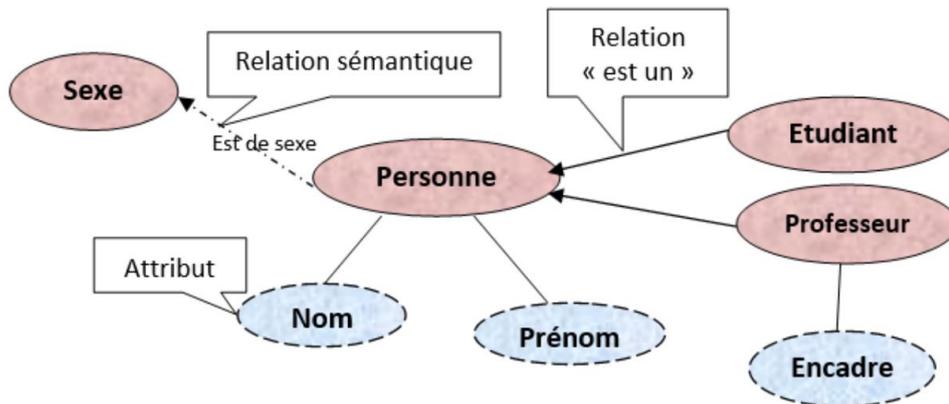


Figure 2.1: Exemple d'une ontologie simple [28].

### 2.3. Classifications des ontologies

Différentes classifications des ontologies ont été représentées dans la figure 2.2.

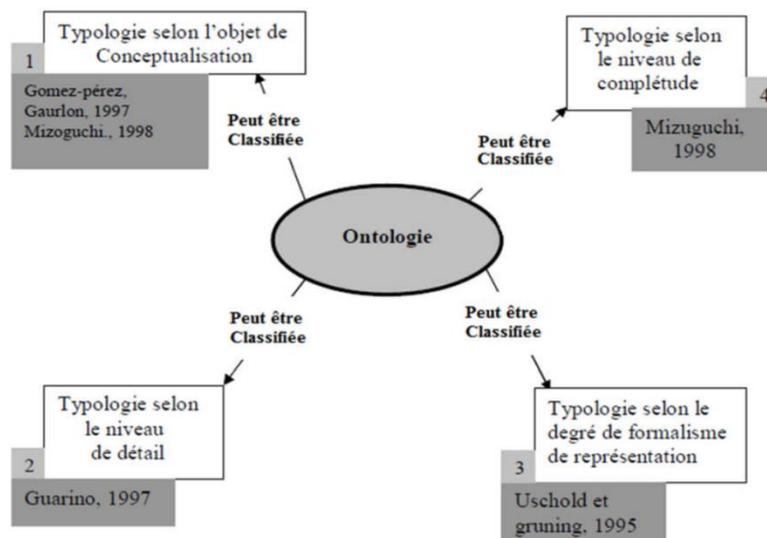


Figure 2.2 : Typologies d'ontologies selon quatre dimensions de classification [29].

### **2.3.1. Selon l'objet de conceptualisation**

#### **2.3.1.1. Ontologie pour la représentation de connaissances**

La création d'une ontologie permet de modéliser la représentation brute des connaissances dans un domaine spécifique. Par exemple, pour formaliser les connaissances liées aux TopicMaps, l'ontologie pourrait englober des concepts comme "Sujet", "Type de Sujet", "Association", "Occurrence", "Type d'Occurrence", et ainsi de suite [30].

#### **2.3.1.2. Ontologies supérieures (Upper or Top Level Ontologie)**

Une ontologie générale est une représentation de concepts applicables à divers domaines. Elle inclut des notions très générales comme l'espace, le temps, la matière, l'objet et l'événement, qui ne sont pas propres à un domaine particulier et nécessitent un consensus communautaire, du moins en théorie. Cette ontologie repose sur la théorie de la dépendance et examine les différentes catégories d'existences dans le monde. Elle englobe des concepts abstraits tels que l'entité, l'événement, l'état, le processus, l'action, le temps, l'espace, la relation et la propriété [30].

#### **2.3.1.3. Ontologie Générique (Generic ontology)**

Également connue sous le nom de noyau ontologique, cette connaissance se caractérise par un niveau d'abstraction inférieur à celui de l'ontologie de haut niveau, tout en restant suffisamment générale pour être réutilisée dans divers domaines. Elle englobe un vocabulaire portant sur les entités, les événements, le temps, l'espace, la causalité, le comportement, les fonctions, et d'autres concepts similaires.

#### **2.3.1.4. Ontologies de domaine**

Cette ontologie formalise les connaissances spécifiques à un domaine particulier et trouve diverses applications au sein de ce domaine. Elle établit les concepts et les relations entre eux, englobant ainsi le vocabulaire, les activités et les théories associés. Une ontologie de domaine représente la connaissance spécifique à ce domaine ou à la tâche en cours. Par exemple, dans le contexte de l'apprentissage en ligne, un domaine pourrait être la formation [30]

#### **2.1.3.5. Ontologies de tâche**

L'ontologie de tâches propose un vocabulaire structuré pour aborder des problèmes spécifiques à des tâches, indépendamment du domaine auquel elles appartiennent. Elle offre un ensemble de termes permettant de décrire de manière générale les méthodes de résolution de différents types de problèmes, en incluant des noms, des verbes et des adjectifs génériques pour décrire les tâches [30].

### 2.1.3.6. Ontologies d'application

Ce sont les ontologies les plus précises. Elles offrent la possibilité de décrire des concepts qui sont liés à la fois à un domaine et à une tâche. Dans cette classification, la notion d'ontologie d'application établit le contexte d'une application qui explique la signification des informations et des services utilisés par une ou plusieurs applications dans un domaine donné [31].

### 2.3.2. Selon le degré de formalisme de la représentation

Il existe quatre catégories dans cette typologie:

#### 2.3.2.1 Hautement informelle

Elle est exprimée en langue naturelle (sémantique ouverte).

**2.3.2.2 Semi-informelle :** Écrite dans un langage naturel organisé et restreint-en suivant un ensemble de règles visant à améliorer la clarté et la lisibilité.

**2.3.2.3 Semi-formelle :** Exprime de manière artificielle et formelle.

**2.3.2.4 Rigoureusement formelle:** Le langage utilisé pour exprimer l'ontologie comprend une sémantique formelle, des théorèmes et des preuves afin de vérifier les caractéristiques telles que la validité et la complétude [32].

### 2.3.3. Selon le niveau de détail

Les ontologies peuvent être classées en fonction du degré de détail utilisé lors de la conception de l'ontologie et de l'objectif opérationnel envisagé. On peut distinguer deux catégories : la granularité fine et la granularité large.

**2.3.3.1 Granularité fine :** Elle représente une ontologie extrêmement approfondie, avec un vocabulaire plus étendu qui permet de décrire de manière détaillée les concepts pertinents d'un domaine ou d'une tâche.

**2.3.3.2 Granularité large :** Elle correspond à un lexique moins approfondi. Les ontologies génériques ont une grande étendue de granularité, car les concepts qu'elles représentent peuvent être enrichis par des concepts plus précis [32].

### 2.3.3 Selon le niveau de complétude

La classification au niveau de complétude est proposée en trois niveaux :

#### 2.3.4.2 Niveau Sémantique

Chaque concept (caractérisés par un terme/libellé) doit se conformer aux quatre principes distincts [32]:

- ✓ Communauté avec l'ancêtre ;
- ✓ Différence (spécification) par rapport à l'ancêtre ;

- ✓ Communauté avec les concepts frères (situés au même niveau) ;
- ✓ Différence par rapport aux concepts frères (sinon il n'aurait pas lieu de le définir).

#### **2.3.4.2 Niveau Référentiel**

Les concepts de base (ou formels) se distinguent par un terme/libellé dont la signification est définie par une série d'objets. Le concept de l'engagement ontologique détermine les éléments du domaine qui peuvent être associés à la notion, en fonction de sa signification officielle. S'ils ont la même extension, deux concepts formels seront identiques [32].

#### **2.3.4.3 Niveau Opérationnel**

Les concepts du domaine opérationnel ou computationnel se distinguent par les opérations qu'ils peuvent être utilisés pour produire des inférences (engagement computationnel). S'ils ont le même potentiel d'inférence, deux concepts opérationnels sont identiques [32].

### **2.4 Avantages de l'ontologie**

- ✓ Réutilisation et partage des connaissances ;
- ✓ Plus stable ;
- ✓ Ontologie au niveau (compréhension par la machine et l'homme) ;
- ✓ Amélioration de l'analyse sémantique ;
- ✓ Extraction d'information efficace ;
- ✓ Flexibilité et extensibilité ;
- ✓ Évolution et Adaptabilité.

### **2.5 Utilisation de l'ontologie**

L'utilisation de l'ontologie implique la création et l'application d'une structure organisée des concepts et des relations entre ces concepts, dans un domaine particulier. Cette ontologie qui inclut ces derniers (les concepts et les relations) couvre du vocabulaire, des activités et des théories spécifiques à ce domaine, afin d'améliorer la compréhension et l'analyse des données.

L'utilisation de l'ontologie, en analyse des opinions, enrichit la compréhension et l'interprétation des textes en intégrant des connaissances structurées et contextuelles spécifiques à un domaine donné. Elle améliore ainsi la précision, la granularité et l'utilité des analyses de sentiments, augmentant ainsi la qualité et la pertinence des analyses. Ce qui offre des insights plus détaillés et exploitables, facilitant la prise de décision et l'amélioration des stratégies dans divers domaines.

Des outils comme Protégé et des plateformes comme GATE sont couramment utilisés pour intégrer des ontologies dans les systèmes d'analyse des sentiments, permettant d'obtenir des insights plus exploitables et précis.

## 2.6 Conclusion

Dans ce chapitre, nous avons présenté les ontologies en commençant par leur définition et en détaillant les différentes typologies selon quatre dimensions de classification : l'objet de conceptualisation, le degré de formalisme de la représentation, le niveau de complétude et le niveau de détail. Nous avons expliqué les éléments spécifiques de chaque type et énuméré les principaux avantages des ontologies. Ensuite, nous avons abordé leurs diverses utilisations, en soulignant que les systèmes d'extraction d'informations s'appuient principalement sur les ontologies pour structurer les connaissances du domaine qu'elles explorent. Notre objectif est d'utiliser cette ontologie pour identifier et extraire diverses opinions. Dans le chapitre suivant, nous présenterons le Bitcoin, la monnaie numérique au cœur de notre étude. Enfin, nous discuterons des recherches déjà menées sur ce sujet.

## **Chapitre 3 :**

Analyse des opinions dans le domaine  
économique

### 3.1 Introduction

L'analyse des opinions dans le domaine économique, notamment en ce qui concerne le Bitcoin, joue un rôle crucial dans la compréhension des tendances et des sentiments des investisseurs, des régulateurs et du grand public. En effet, le Bitcoin est une cryptomonnaie volatile qui suscite des opinions diverses quant à son potentiel d'investissement, sa légitimité en tant que monnaie et son impact sur les marchés financiers traditionnels. L'analyse des opinions vise à extraire à partir de sources telles que les médias sociaux, les actualités financières et les forums en ligne, les sentiments et les attitudes des différents acteurs à l'égard du Bitcoin, ce qui peut éclairer les décisions d'investissement, les politiques réglementaires et les prévisions de marché.

Dans ce chapitre, nous définissons le Bitcoin et explorons son fonctionnement, ses caractéristiques, ainsi que ses avantages et inconvénients.

### 3.2 Bitcoin

#### 3.2.1. Définition

Le Bitcoin, première monnaie virtuelle et électronique, a déclenché une véritable révolution dans le domaine des opérations financières. Utilisant la cryptographie pour vérifier les transactions entre les participants du réseau, il permet l'achat de biens et de services en ligne. Le terme "Bitcoin" est une contraction de "bit" (unité de mesure binaire) et "coin" (pièce de monnaie). Contrairement aux monnaies traditionnelles, il n'est pas nécessaire de créer un compte Bitcoin ; les Bitcoins peuvent être détenus comme des pièces de monnaie dans un portefeuille.

La particularité du Bitcoin réside dans son absence de banque centrale ou d'institution financière centralisée pour le contrôler, s'appuyant plutôt sur un réseau décentralisé de pair-à-pair sur Internet [33]. Le Bitcoin n'est pas matérialisé et n'est soutenu par aucune autorité gouvernementale. Il est créé par des individus et des entreprises à travers le monde qui utilisent des logiciels pour résoudre des problèmes mathématiques complexes. Actuellement, le Bitcoin est le leader en matière de capitalisation boursière [33].

En 2010, une des premières transactions en Bitcoin a été réalisée par un programmeur nommé Laszlo Hanyecz, qui a acheté deux pizzas pour 10 000 Bitcoins, à une époque où cette monnaie n'avait pratiquement aucune valeur réelle. Bitcoin a commencé à gagner en popularité très lentement, principalement utilisé à ses débuts par des passionnés d'informatique. L'objectif principal de Bitcoin est de faciliter les paiements en dehors du système bancaire traditionnel [34].

### 3.3 Fonctionnement de Bitcoin

Le fonctionnement de Bitcoin repose sur plusieurs éléments clés, la figure 3.1 montre le fonctionnement de Bitcoin.

#### 3.3.1 Blockchain

La Blockchain, technologie de registre distribué, est essentielle au fonctionnement de Bitcoin, consignait toutes les transactions dans une succession de blocs. Chaque bloc contient l'historique des transactions récentes et une référence au bloc précédent, créant une chaîne ininterrompue de blocs. Cette chaîne est maintenue par un réseau de nœuds, ou des ordinateurs qui valident et enregistrent les transactions sans l'intervention d'une autorité centrale. La blockchain est entièrement publique et transparente, permettant à quiconque de consulter l'intégralité de l'historique des transactions de Bitcoins [35].

#### 3.3.2 Processus de transaction

Une opération de transaction de Bitcoins implique le déplacement de valeur entre des portefeuilles Bitcoin, avec cet échange de données étant enregistré dans la blockchain. Chaque portefeuille Bitcoin inclut une donnée sensible, la clé privée, qui sert à authentifier les transactions, assurant ainsi, par une preuve mathématique, l'origine légitime de la transaction du propriétaire du portefeuille. Cette authentification empêche toute altération de la transaction post-émission qui est diffusée entre les utilisateurs et typiquement validées par le réseau dans les dix minutes suivantes, grâce à un processus appelé minage [35].

#### 3.3.3 Nœuds

Le réseau Bitcoin est composé de multiples nœuds, lesquels sont des ordinateurs exécutant le logiciel Bitcoin. Ces nœuds ont pour rôle de valider les transactions, de créer des blocs et de maintenir la blockchain. Les nœuds sont répartis dans le monde entier et sont administrés par des bénévoles.

Il existe plusieurs types de nœuds, notamment les nœuds complets et les nœuds mineurs. Les nœuds complets conservent une copie intégrale de la blockchain et vérifient chaque transaction conformément aux règles du protocole Bitcoin. Les mineurs, en revanche, sont des nœuds spécialisés et qui s'occupent de la création de nouveaux blocs en résolvant des problèmes cryptographiques complexes [36].

#### 3.3.4 Proof-of-work (PoW)

Le Proof of Work (Preuve de Travail) est un mécanisme crucial pour établir la confiance dans un environnement décentralisé tel que celui du Bitcoin. Pour qu'une transaction soit validée et ajoutée à la blockchain, des mineurs utilisent des ordinateurs

puissants pour résoudre des énigmes cryptographiques complexes. Le premier mineur à résoudre l'énigme diffuse sa solution au réseau. Après vérification, le bloc de transactions est enregistré définitivement dans la blockchain. Ce processus nécessite une quantité significative de puissance de calcul, ce qui rend économiquement inviable toute tentative de manipulation du système.

Le Proof of Work garantit la sécurité du réseau Bitcoin en rendant extrêmement coûteuse et difficile la modification de la blockchain. Pour altérer une transaction antérieure, un attaquant devrait refaire le travail de minage pour le bloc contenant cette transaction ainsi que pour tous les blocs suivants. Cette tâche nécessiterait une énorme quantité d'énergie et de puissance de calcul, rendant une telle attaque pratiquement impossible. De plus, étant donné que le réseau Bitcoin est distribué sur des milliers de nœuds, une attaque réussie exigerait que la majorité de ces nœuds soient compromis, ce qui renforce encore la sécurité du réseau [36].

### 3.3.5 Minage

Le minage est le processus par lequel les transactions sont vérifiées et ajoutées à la blockchain, tout en créant de nouveaux Bitcoins. Les mineurs reçoivent une récompense en Bitcoins nouvellement créés pour leur travail, appelée récompense de bloc. Cette récompense est réduite de moitié à des intervalles réguliers lors d'un événement connu sous le nom de halving. Le minage sert donc à la fois de mécanisme de sécurisation du réseau et de méthode de distribution de nouveaux Bitcoins dans le système. C'est une compétition ouverte, incitant les mineurs à améliorer constamment l'efficacité et la sécurité du réseau [36].

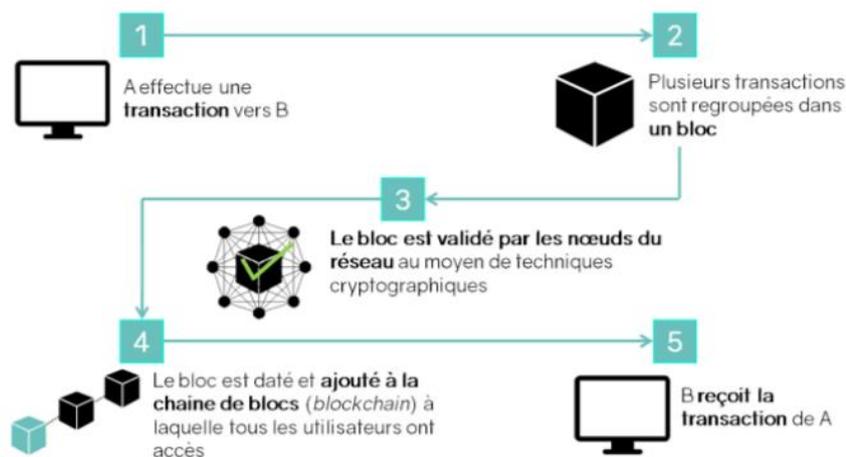


Figure 3.1 : Un exemple de fonctionnement de Bitcoin [37].

### 3.4 Caractéristiques de Bitcoin

Les principales caractéristiques du Bitcoin sont les suivantes :

#### 3.4.1 Décentralisation

Bitcoin est un système décentralisé, exempt de contrôle par une autorité centrale telle qu'une banque ou un gouvernement. Le réseau Bitcoin se compose de nombreux nœuds qui travaillent ensemble pour traiter et valider les transactions. Cette structure collective empêche toute manipulation du système ou gel des fonds des utilisateurs. La technologie de la blockchain permet cette décentralisation [38].

#### 3.4.2 Quantité limitée

Le nombre total de Bitcoins est plafonné à 21 millions, chaque unité étant divisible jusqu'à la huitième décimale. La génération de nouveaux Bitcoins par le processus de minage est réduite de moitié tous les quatre ans. Contrairement aux monnaies fiduciaires traditionnelles, qui peuvent être émises sans restriction par les banques centrales, cette limitation intégrée est une caractéristique essentielle de Bitcoin, destinée à maintenir sa valeur sur le long terme [38].

#### 3.4.3 Transparence

Chaque transaction en Bitcoin est publique et enregistrée dans la blockchain, qui agit comme un registre décentralisé. Les protocoles, algorithmes et données étant accessibles à tous, le réseau peut être examiné par n'importe qui. Malgré cela, les utilisateurs peuvent rendre leurs activités plus difficiles à suivre en utilisant des méthodes comme les transactions en chaîne ou les portefeuilles jetables [38].

#### 3.4.4 Anonymat relatif

Les utilisateurs peuvent posséder plusieurs adresses Bitcoin sans qu'elles soient associées à des informations d'identification personnelle. Toutefois, les transactions peuvent toujours être suivies via la blockchain. Ainsi, le Bitcoin n'est pas entièrement anonyme, mais plutôt pseudonyme. L'anonymat dépend également des habitudes des utilisateurs et des pratiques des prestataires de services [38].

#### 3.4.5 Facilité de mise en place

La configuration d'une adresse Bitcoin est simple et rapide, sans nécessiter de formalités administratives ni de frais. Les transactions sont presque instantanées et les frais sont très faibles, ce qui les distingue des virements bancaires traditionnels, souvent lents et coûteux. Cette facilité d'utilisation rend Bitcoin particulièrement avantageux pour les paiements transfrontaliers et les micro-paiements [38].

### 3.5 Avantages de Bitcoin

Le Bitcoin, en tant que la première et la plus célèbre des monnaies numériques, présente un certain nombre d'avantages et d'inconvénients [39,40] :

- Transferts de fonds faciles entre deux parties sans l'intervention d'un intermédiaire tel qu'une banque.
- Alternative moins coûteuse aux transactions en ligne traditionnelles.
- Offre un haut niveau d'anonymat et de sécurité lors des transactions.
- Les systèmes modernes de crypto-monnaie incluent un "portefeuille" ou une adresse de compte utilisateur accessible uniquement par une clé publique et une clé privée connue uniquement du propriétaire du portefeuille.
- Système décentralisé offrant une certaine résilience face aux manipulations.
- Possibilité de réaliser des transferts d'argent quasi instantanés.
- Opportunité d'investissement et de gains rapides pour les particuliers.
- Transparence : La technologie blockchain du Bitcoin assure la transparence des transactions.

### 3.6 Inconvénients de Bitcoin

Bitcoin présente également certains inconvénients [38,39,40,41] :

- Les crypto-monnaies ne sont pas acceptées partout et n'ont qu'une valeur limitée ailleurs.
- Permet des activités illégales comme le blanchiment d'argent.
- Valeur très volatile, le Bitcoin ayant connu des flambées et effondrements rapides.
- Aucune garantie de valeur minimale, exposant les utilisateurs à des pertes potentielles en cas de chute de la valeur.
- Risque de perte de portefeuilles électroniques en cas de défaillance matérielle ou de corruption des données.
- Acceptation limitée, certains pays refusant encore les transactions en Bitcoin.

Ainsi, le Bitcoin offre des avantages en termes de rapidité, de réduction des coûts et de sécurité, mais présente également des inconvénients majeurs tels que son adoption limitée, sa volatilité, les risques de perte et de sécurité, ainsi que l'absence de protection pour les utilisateurs.

### 3.7 Travaux antérieurs

#### 3.7.1 The Relationship Between Social Media Sentiment and Bitcoin Price Volatility

Ce travail se concentre sur la corrélation entre le sentiment exprimé sur les médias sociaux et la volatilité du prix du Bitcoin, en utilisant Python. Il utilise des techniques d'analyse de texte et de traitement du langage naturel (NLP) pour extraire et interpréter le sentiment des publications liées au Bitcoin sur les plateformes de médias sociaux. Parallèlement, il emploie des méthodes d'analyse financière et de données, notamment basées sur l'apprentissage automatique, pour mesurer et quantifier la volatilité du prix du Bitcoin sur la période étudiée. Une analyse statistique est ensuite menée pour rechercher des corrélations entre le sentiment des médias sociaux et la volatilité des prix, avec pour objectif potentiel de développer des modèles prédictifs.

L'objectif principal de ce travail est d'explorer comment le sentiment exprimé dans les médias sociaux peut influencer le marché du Bitcoin. Ces recherches visent à fournir des insights essentiels pour les investisseurs et les acteurs du secteur des crypto-monnaies, en offrant une perspective précieuse sur cette dynamique [42].

Les étapes de cette étude sont résumées dans la figure 3.2.

- **Création d'un Contexte Théorique** : Commence par établir un cadre théorique comprenant les termes et concepts pertinents pour la compréhension du sujet général de l'étude ;
- **Examen des Études Empiriques** : Révision des recherches empiriques liées au sujet de recherche pour conceptualiser davantage la problématique ;
- **Identification des Données Nécessaires** : Définition du type de données nécessaires pour atteindre l'objectif de le travail ;
- **Collecte de Données sur Twitter** : Utilisation d'un outil Python pour collecter des données pertinentes sur Twitter ;
- **Nettoyage et Prétraitement des Données** : Utilisation d'outils Python pour nettoyer et prétraiter les données brutes, incluant des techniques de nettoyage de texte et de normalisation ;
- **Analyse des Données** : Utilisation de trois modèles différents : analyse du sentiment, modélisation des sujets et analyse du sentiment basée sur les aspects, pour analyser les données prétraitées ;
- **Discussion des Résultats** : Débat sur les résultats obtenus à partir des modèles et des outils d'analyse utilisés.

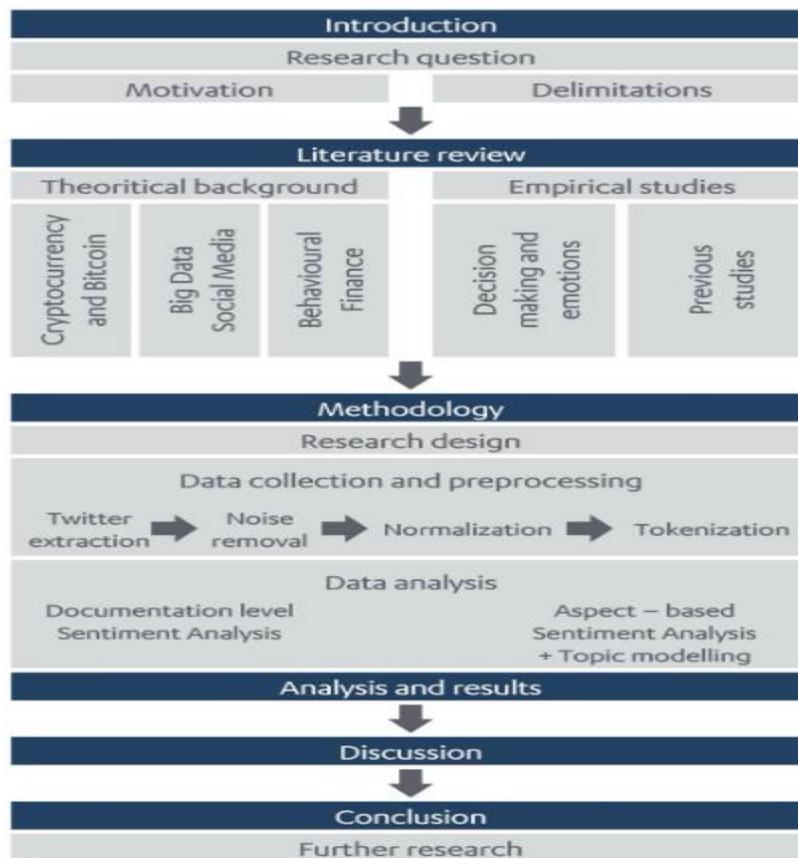


Figure 3.2 : Schéma proposé de cette étude [42].

### 3.7.2 Sentiment-Driven Cryptocurrency Price Prediction: A Machine Learning Approach Utilizing Historical Data and Social Media

Cette étude se concentre sur le développement d'un modèle de prédiction des prix des crypto monnaies en intégrant le sentiment des médias sociaux avec les données historiques de prix et d'autres variables pertinentes. En utilisant des techniques d'analyse de sentiment pour évaluer le ton général des discussions sur les crypto monnaies sur les réseaux sociaux, le modèle vise à capturer les tendances émotionnelles du marché et à les exploiter pour améliorer la précision des prédictions de prix. La méthodologie comprend la collecte et le traitement des données historiques et des données de sentiment, ainsi que l'utilisation d'algorithmes d'apprentissage automatique pour entraîner et évaluer le modèle prédictif. Les résultats de l'étude fournissent des insights sur l'efficacité de cette approche hybride, offrant ainsi des perspectives précieuses pour les investisseurs et les analystes intéressés par la prédiction des prix dans le domaine des cryptomonnaies.

L'objectif est de créer un modèle prédictif des prix des cryptomonnaies en utilisant à la fois les données historiques de prix et le sentiment exprimé sur les réseaux sociaux. L'idée est d'améliorer la capacité de prédiction en intégrant ces deux sources d'information, afin d'aider

les investisseurs et les analystes à prendre des décisions plus informées dans le domaine des crypto monnaies.

La figure 3.3 ci-dessous explique la méthodologie utilisée dans cette recherche. Ils ont calculé le sentiment des Tweets à l'aide de VADAR et de Twitter-RoBERTa, et ils ont entraîné cinq modèles de classification, à savoir Naïve Bayes, K-NearestNeighbours, Support Vector Machine, LogisticRegression, et XGBoost, ainsi qu'un Multi Modal Fusion. XGBoost, ainsi qu'un modèle de fusion multimodale [43].

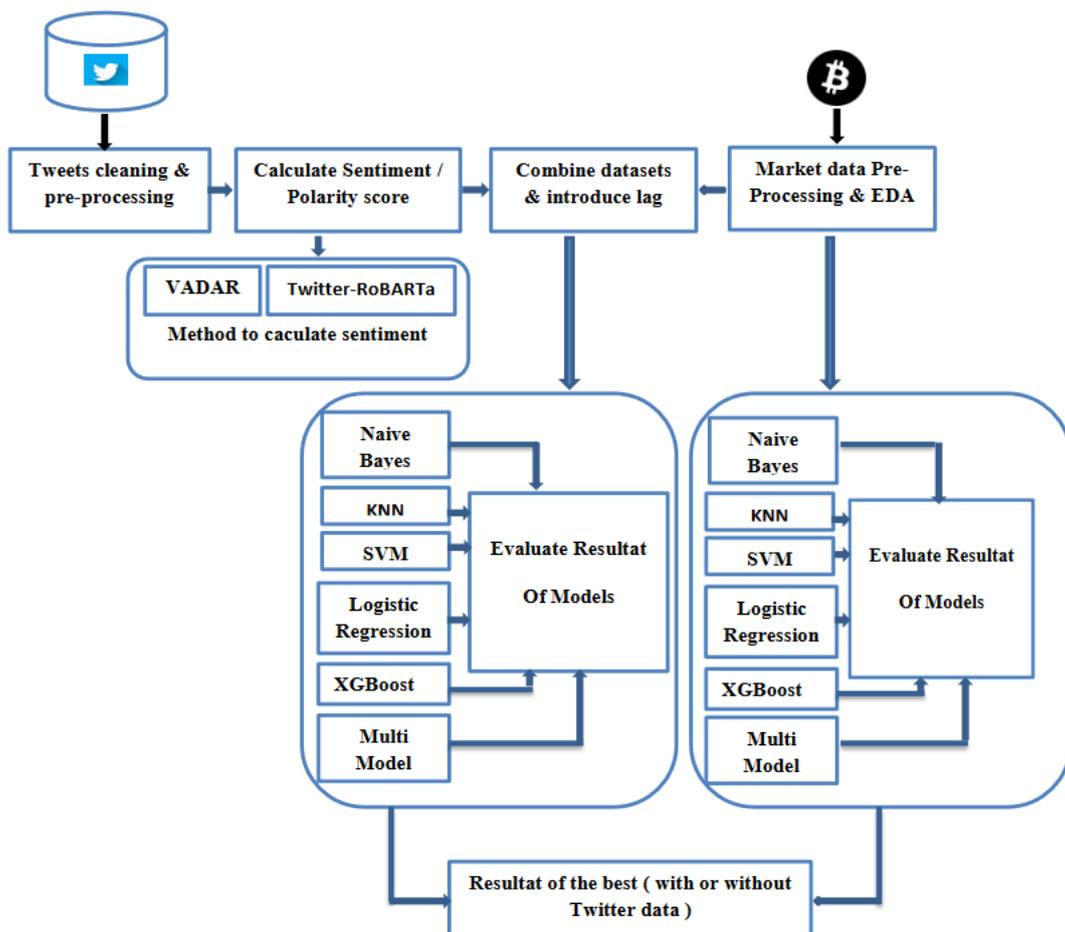


Figure 3.3 : Méthodologie utilisée dans cette recherche [43].

### 3.8 Conclusion

Dans ce chapitre, nous avons présenté la définition du Bitcoin ainsi que son fonctionnement, qui repose sur plusieurs éléments clés : la Blockchain, le processus de transaction, les nœuds, le Proof-of-Work (PoW) et le minage. Nous avons également discuté des principales caractéristiques du Bitcoin, mettant en avant ses avantages tels que la facilité de transfert, l'anonymat relatif des transactions et l'absence d'intermédiaires financiers traditionnels. En revanche, nous avons identifié ses inconvénients, notamment sa volatilité extrême, les risques de sécurité et son acceptation limitée par certains pays.

Malgré ces défis, le Bitcoin continue de susciter un intérêt croissant en tant qu'actif financier alternatif et comme moteur de l'innovation dans les technologies financières. Enfin, nous avons abordé quelques études connexes.

Dans notre étude, nous avons utilisé le Bitcoin pour extraire les facteurs qui affectent la valeur du Bitcoin et les opinions des gens sur le Bitcoin. Dans le chapitre suivant nous proposons et expliquons notre approche avec ses étapes.

Chapitre 4 :  
Conception d'un système de  
détection sémantique des  
opinions

## 4.1. Introduction

Dans le chapitre précédent, nous avons défini le Bitcoin et exposé son fonctionnement en détail. Ensuite, nous avons discuté de ses caractéristiques, cité ses avantages et inconvénients, et mentionné les travaux liés.

Dans ce chapitre, nous présenterons une architecture générale de notre approche proposée pour un système d'analyse sémantique des opinions sur la plateforme Facebook, basée sur une ontologie relative à la monnaie numérique Bitcoin. Nous définirons également notre approche ainsi que l'architecture et les étapes de cette proposition.

## 4.2. Notre Approche Proposée

Actuellement, et à travers le monde entier, les réseaux sociaux sont considérés comme la source la plus visitée pour s'informer en tout ce qui concerne les technologies modernes, tel Bitcoin.

Bitcoin est la crypto-monnaie la plus importante avec la plus grande capitalisation boursière. De plus, il s'agit d'une monnaie numérique à laquelle les utilisateurs ne peuvent accéder qu'en ligne. Ainsi, les plateformes en ligne jouent un rôle crucial dans la transition de l'information aux individus sur Bitcoin et sur la manière de son utilisation. Les individus font principalement appel aux réseaux sociaux pour prendre des décisions d'achat, y compris l'achat ou l'investissement dans le Bitcoin, c'est pourquoi nous avons choisi les réseaux sociaux particulièrement Facebook car il regroupe toutes les catégories de notre société.

Dans notre étude, nous avons étudié des articles de recherche dans le domaine pour extraire des informations sur les facteurs qui affectent le Bitcoin, puis nous allons les classés en trois catégories : les facteurs positifs, les facteurs négatifs et les facteurs neutres.

### ❖ Facteur positif

Nous avons identifié plusieurs facteurs positifs qui ont un impact sur l'augmentation du Bitcoin, parmi lesquels nous en citons quelques-uns : la demande croissante, l'adaptation des institutions, l'inflation et l'instabilité économique, une couverture médiatique accrue et d'autres éléments.

### ❖ Facteur négatif

La dépréciation du Bitcoin est influencée par plusieurs facteurs, dont certains sont mentionnés ci-dessous : une volatilité élevée, des crises économiques, telles que les guerres, des taux d'intérêt élevés, la concurrence des autres crypto-monnaies, la difficulté de l'utiliser comme monnaie et d'autres facteurs.

#### ❖ Facteur neutre

Il y a des éléments neutres, dont certains sont mentionnés ci-dessous : évaluation de la concurrence, stabilité, actualité médiatique.

L'objectif de l'extraction des éléments qui influencent la valeur du Bitcoin est de mieux appréhender le marché et de prédire les tendances futures, de renforcer la confiance des individus dans le Bitcoin, d'encourager son utilisation, d'aider à étendre son utilisation dans différents domaines, d'améliorer les performances des échanges et d'autres plateformes, et d'aider davantage de personnes à mieux comprendre cette monnaie. De plus, cela permet de comprendre les risques liés à l'investissement dans le Bitcoin et de protéger les consommateurs contre la fraude.

Aussi, nous nous focalisons sur l'analyse des opinions sur Bitcoin à travers des publications et des commentaires sur Facebook sur le prix de Bitcoin et leur niveau de satisfaction et risques envers ce dernier. Par ces avis, il est envisageable de :

- déterminer à quel point cette monnaie est répandue ;
- déterminer si les individus sont optimistes ou pessimistes quant à son avenir et de mieux appréhender leurs besoins ;
- mesurer la confiance des gens dans cette monnaie, leur niveau de satisfaction à son égard et leurs attentes futures ;
- permettre aux développeurs de concevoir de nouvelles technologies pour améliorer l'efficacité du marché ;
- faciliter les transactions et sensibiliser aux risques associés à l'investissement dans cette monnaie, offre également la possibilité de reconnaître son influence sur l'économie et la communauté.

### 4.3. Architecture de modèle proposé

L'architecture ci-dessous représente la conception détaillée de notre système (figure 4.1). Le système proposé se compose de plusieurs étapes :

- ✓ Collection des données ;
- ✓ Prétraitement ;
- ✓ Création d'ontologie de domaine ;
- ✓ Analyse des informations ;
- ✓ Machine Learning.

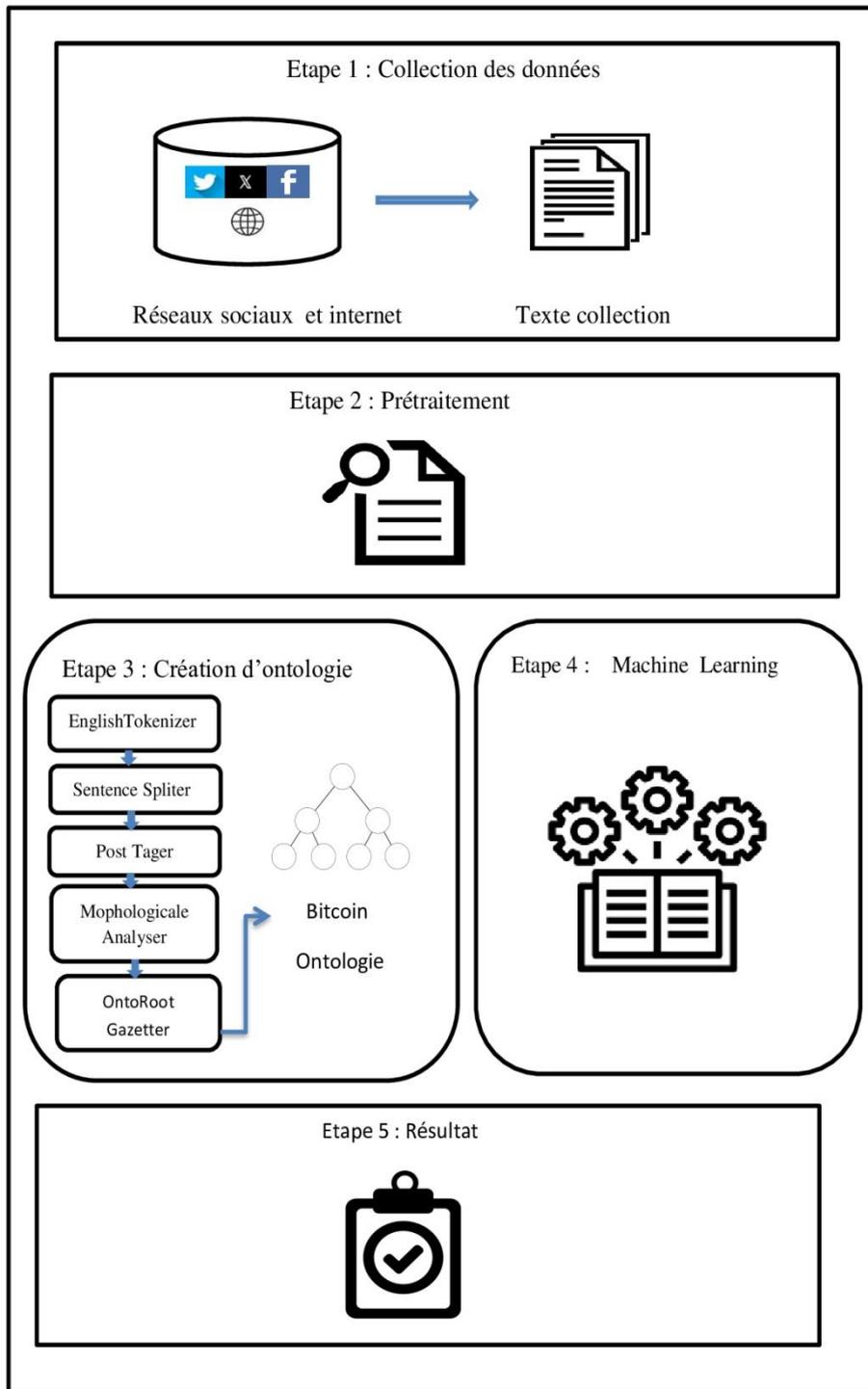


Figure 4.1: Architecture générale du système proposé.

## 4.4 Étapes de l'approche proposée

### 4.4.1 Étape de collection des données

Nous obtenons des informations sur Internet et sur les réseaux sociaux (Facebook). Nous essayons de collecter manuellement les informations relatives aux opinions des gens sur le Bitcoin à partir de Facebook et les facteurs qui affectent la valeur du Bitcoin sont collectés à partir d'internet, tels que les facteurs réglementaires, économiques, environnements et autres.

### 4.4.2 Prétraitement

Nous avons nettoyé et normalisé les données textuelles et les commentaires pour les rendre exploitables concernant le Bitcoin, puis ils seront traités dans la prochaine étape.

### 4.4.3 Étape de création de l'ontologie

La construction de l'ontologie est l'objet principal de cette étude. Dans ce processus, nous avons opté pour une approche descendante. Dans un premier temps, nous avons repéré les concepts abstraits, puis nous les avons convertis en concepts plus précis dans notre ontologie, connue sous le nom de Bitcoin ontologie, qui représente les connaissances essentielles de notre travail. Cette ontologie a été élaborée de manière manuelle, puis mise en place en utilisant l'outil Protégé au format OWL.

Comme mentionné précédemment, le processus de développement manuel d'ontologies comprend les étapes suivantes :

- Déterminer le domaine et la portée de l'ontologie ;
- Envisager de réutiliser les ontologies existantes ;
- Énumérer les termes importants de l'ontologie ;
- Définir les classes et la hiérarchie des classes ;
- Définir les propriétés des classes (slots) ;
- Définir les facettes des fentes ;
- Créer des instances.

Tableau 4.1 : Classes, sous-classes et sous sous-classes de l'ontologie.

Classes	Sous – classes	Sous sousclasses
<b>Facteur positif</b>	Adaptation	Grand_public Institutionnelle
	Demande	
	Reglementaire_positif	Acceptation
	Technologie_positif	Couverture_mondiale Halving Sécurité_de_réseau
	Economique_positif	Inflation Instabilité_économique Performance_des_marché_f Rareté Speculation_accrue
<b>Facteur négatif</b>	Concurrence	Développement_d_c_m
	Environnement	Consommation_d'énergie
	Reglementaire_négatif	Interdiction Manque_d_c_juridique Surveillance_c_p_gv
	Technologie_négatif	Piratage
	Volatilité	Déficience_u_c_monnaie Risque_p_i_et_u Volatilité_d_prix
	économique_négatif	Équilibre_avec_m_trd Les_crisis_éco Manipulation_d_marché Retirer_s_entr Taux_d'interet_e
<b>Facteur neutre</b>	Actualité_médiatique Évaluation_de_concurrence Stabilité	
<b>Opinion positif</b>	le meilleur. des chiffres très solides. la monnaie la plus forte. grande réserve d'argent. maximiser la liberté individuelle. niveau élevé de protection de la vie privée. révolution financière. symbole de l'évaluation future. véritable joyau.	
<b>Opinion négatif</b>	bulle spéculative. consommation d'énergie. forte volatilité. illégal. instable. non garanti. peu sûr.	

	risqué. Sujet à la manipulation. Virtuel .	
<b>Opinion neutre</b>	peut s'agir d'un échec total. peut s'agir d'une grande réussite. simple moyen d'investir. significatif technologie émergente. une nouvelle expérience économique. un avenir incertain.	

**4.4.4 Étape d'analyse d'un texte**

**4.4.4.1 Tokeniser**

Le Tokeniser divise le texte en jetons très simples tels que des chiffres, des signes de ponctuation et des mots de différents types. Par exemple, nous distinguons les mots en Majuscule et en Minuscule, et entre certains types de ponctuation, etc. En ajoutant une annotation "Jeton" à chacun, il n'a pas besoin d'être modifié pour différentes applications ou types de texte. L'objectif est de limiter le travail du Tokeniser pour maximiser l'efficacité et permettre une plus grande flexibilité en mettant la charge sur les règles de grammaire, qui sont plus adaptables [44].

**Les étapes de l'analyse des informations**

1)

Phrase non analyser  
 I like working and making money.

2)

Décalage des caractères  
 I like working and making money.  
 0..2....7.....15.....25.....|

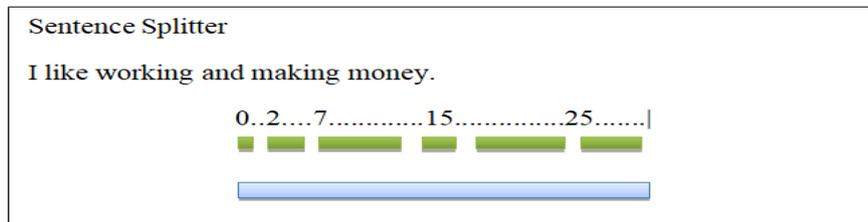
3)

Tokeniser  
 I like working and making money.  
 0..2....7.....15.....25.....|  


**Figure 4.2:** Exemple d'un tokeniser.

#### 4.4.4.2 Sentence Splitter

Le séparateur de phrases est une cascade de transducteurs à état fini qui segmente le texte en phrases. Ce module est requis pour le tagueur. Le séparateur utilise une liste d'abréviations du répertoire géographique pour aider à distinguer les points de marquage des phrases des autres types [44].

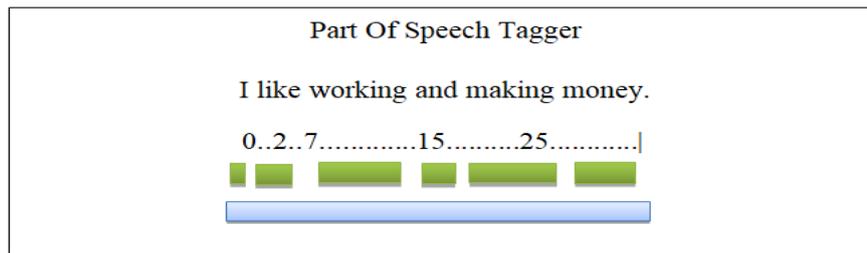


**Figure 4.3:** Exemple d'un Sentence Splitter.

#### 4.4.4.3 Part Of Speech Tagger

Le tagueur est une version modifiée du marqueur Brill, qui produit une étiquette de partie de discours en tant qu'annotation sur chaque mot ou symbole. Le tagueur utilise un lexique et un ensemble de règles par défaut (résultat d'un apprentissage sur un large corpus tiré du Wall Street Journal). Ces deux éléments peuvent être modifiés manuellement si nécessaire. Deux lexiques supplémentaires existent ; un pour les textes tout en majuscules et un pour les textes tout en minuscules. Pour les utiliser, le lexique par défaut doit être remplacé par le lexique approprié au moment du chargement. L'ensemble de règles par défaut doit toujours être utilisé dans ce cas [44]. Le tagueur ANNIE Part-of-Speech requiert certains paramètres :

- ❖ Encoding
- ❖ Lexicon URL
- ❖ Rules URL
- ❖ Document
- ❖ Input AS Name
- ❖ Output AS Name
- ❖ BaseTokenAnnotationType
- ❖ Base Sentence Annotation Type
- ❖ Output AnnotationType
- ❖ PosTag All Tokens
- ❖ Fail On Missing Input Annotations.



**Figure 4.4:** Exemple d'un Part Of Speech Tagger.

**I** : Pronoun (PRP).

**like** : Verb (VBP).

**working** : Gerund or Present Participle (VBG).

**and** : Coordinating Conjunction (CC).

**making** : Gerund or Present Participle (VBG).

**money** : Noun (NN).

#### 4.4.4.4 Analyseur morphologique

Un analyseur morphologique est le module qui réduit les morphes à leur forme canonique. Chaque mot peut avoir une variété de formes et chaque forme peut décrire le même concept, mais d'un point de vue différent.

Prenons par exemple le verbe « courir » et sa forme conjuguée « couraient ». L'analyseur morphologique traitera le jeton « couraient » en le décomposant en sa racine "root=cour" et son suffixe "suffix=aient". Cette transformation est essentielle dans le pipeline, car le répertoire topographique Onto Root s'appuie sur la racine des mots pour effectuer des recherches précises dans l'ontologie [44].

#### 4.4.4.5 OntoRootGazetteer

OntoRootGazetteer (représenté dans la figure 4.2) est un type de répertoire gazeux créé dynamiquement qui, en combinaison avec quelques autres ressources génériques GATE, est capable de produire des annotations basées sur l'ontologie. Avec quelques autres ressources génériques GATE, capable de produire des annotations basées sur l'ontologie et sur le contenu donné par rapport à l'ontologie donnée. Cette gazetteer fait partie du plugin " Gazetteer OntologyBased " qui a été développée dans le cadre du projet TAO (Transitioning Applications to Ontologies)[44].

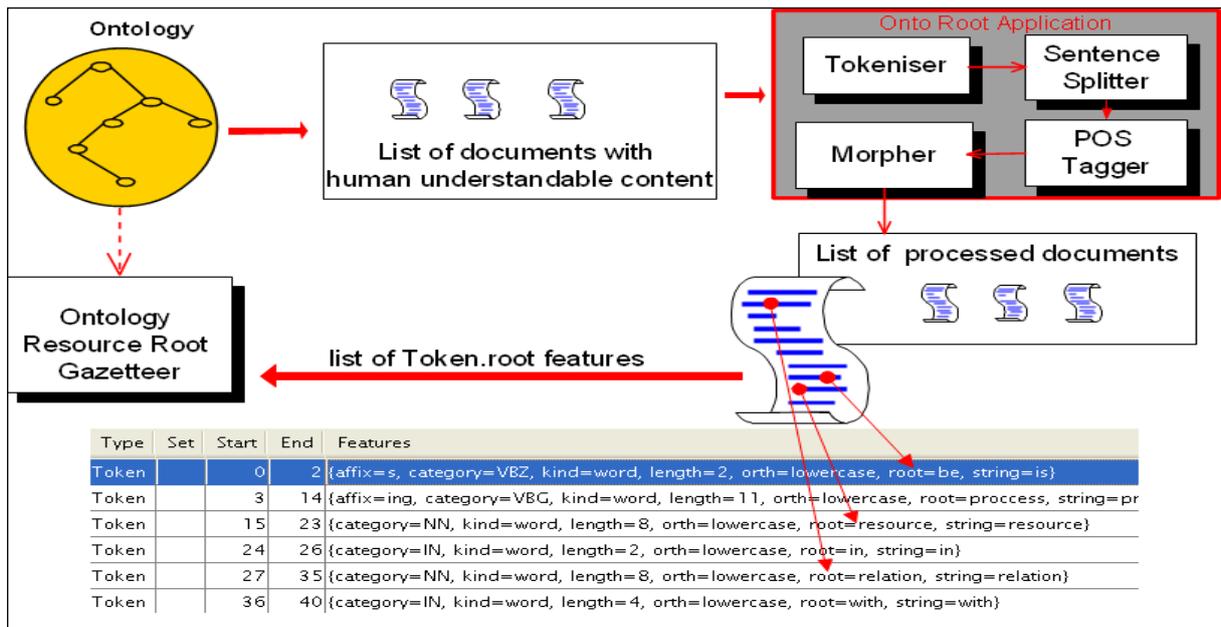


Figure 4.5: Building Ontology ResourceRoot (OntoRoot) Gazetteer from the Ontology.

#### 4.4.5 Machine Learning

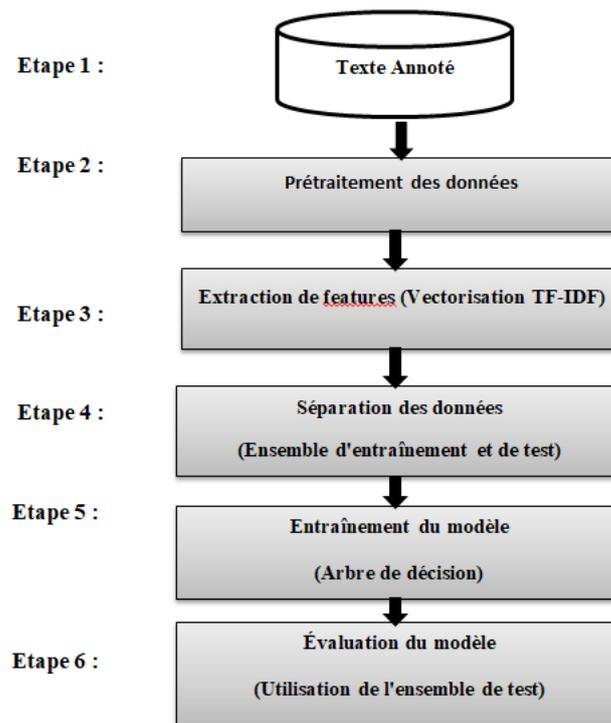
La machine Learning est une discipline de l'intelligence artificielle qui permet aux systèmes informatiques d'apprendre et de s'améliorer automatiquement à partir de l'expérience. En utilisant des algorithmes et des modèles mathématiques, elle analyse les données pour reconnaître des patterns et prendre des décisions sans être explicitement programmée. Les applications de la machine Learning sont vastes, allant de la reconnaissance vocale à la recommandation de produits en ligne, en passant par la détection de fraudes et la conduite autonome. Ce domaine connaît un développement rapide grâce aux progrès technologiques et à la disponibilité croissante de données massives, ouvrant de nouvelles perspectives dans de nombreux secteurs industriels et scientifiques.

#### 4.4.6 Conception d'un modèle basé sur machine Learning

##### 4.4.6.1 Notre proposition

Nous avons récupéré les données qui sont les opinions sur Bitcoin et nous avons traité avec la machine Learning pour comparer entre les deux méthodes qui sont la méthode basée sur l'ontologie et la machine Learning.

Le schéma suivant représente le fonctionnement de la machine Learning de notre travail (Figure 4.6).



**Figure 4.6:** Schéma de fonctionnement de la machine Learning.

Nous avons décrit ces étapes afin de faciliter la compréhension du fonctionnement de la machine Learning :

### 1) Texte annoté

Nous avons chargé les opinions textuelles sur Bitcoin ainsi que les étiquettes correspondantes, indiquant le sentiment associé à chaque opinion (positif, négatif ou neutre).

### 2) Prétraitement des données

Nous avons vérifié que chaque opinion textuelle possède une étiquette correspondante afin de garantir la cohérence des données. De plus, les étiquettes textuelles ("positif", "neutre", "négatif") ont été converties en valeurs numériques (1 pour positif, 0 pour neutre, -1 pour négatif). Nous avons effectué le nettoyage et la tokenisation des données afin de les préparer pour la prochaine étape.

### 3) Extraction de features (ou extraction de caractéristiques)

Utilisation de la technique de vectorisation TF-IDF pour transformer les textes en matrices TF-IDF afin d'obtenir une représentation numérique.

#### 4) Séparation des données

Les données sont divisées en deux ensembles principaux : un ensemble d'entraînement 80% et un ensemble de test 20%. L'ensemble d'entraînement est utilisé pour construire le modèle, tandis que l'ensemble de test est utilisé pour évaluer ses performances.

#### 5) Entraînement du modèle

Le modèle d'arbre de décision a été initialisé et entraîné en utilisant les données d'entraînement.

#### 6) Évaluation du modèle

Le modèle est évalué sur l'ensemble de test pour évaluer ses performances en termes de précision, de rappel et de F1-score.

#### 4.4.6.2 Algorithme de l'arbre de décision

Les arbres de décision sont des méthodes d'apprentissage non paramétriques employées pour des problèmes de classification et de régression. Leur objectif est de construire un modèle capable de prédire les valeurs de la variable cible en utilisant une série de règles de décision dérivées des données d'entraînement [45]. La figure 4.4, représente un exemple de l'arbre de décision.

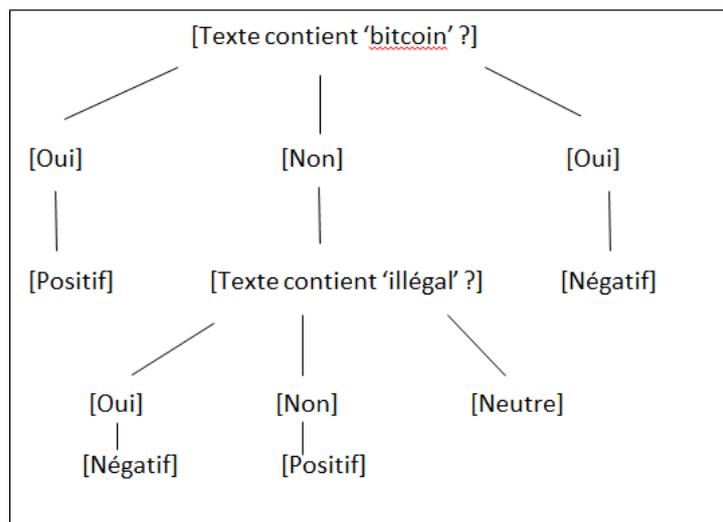


Figure 4.4: un exemple de l'arbre de décision.

#### 4.4.6.3 Arbre de DecisionTreeClassifier

L'arbre de décision acquiert des connaissances sous la forme d'un arbre, qui peut également être réécrit sous la forme d'un ensemble de règles discrètes pour en faciliter la compréhension. Le principal avantage du classificateur par arbre de décision est sa capacité à utiliser différents sous-ensembles de caractéristiques et règles de décision à différents stades

de la classification [46]. Dans notre travail, L'arbre de Décision Tree Classifier est utilisé pour entraîner un modèle de classification des sentiments à partir des données textuelles d'entraînement, puis pour évaluer sa capacité à prédire avec précision les sentiments des nouveaux textes de l'ensemble de test. L'idée principale derrière l'algorithme de l'arbre de décision est de diviser récursivement l'ensemble d'entraînement en sous-ensembles plus petits en fonction des caractéristiques des données, en maximisant la séparation entre les différentes classes de sentiments. Une fois l'arbre de décision est construit, il est capable de prédire avec précision les sentiments des nouveaux textes de l'ensemble de test en les faisant descendre l'arbre selon les règles de décision apprises lors de l'entraînement. En évaluant le modèle sur l'ensemble de test, nous pouvons mesurer sa performance en termes de précision, de rappel, de F1-score et d'autres métriques, ce qui nous permet de comprendre à quel point le modèle est capable de généraliser et de prédire correctement les classes des données de test.

## **4.5 Conclusion**

Dans ce chapitre nous avons détaillé les étapes de notre approche d'un système de détection sémantique des opinions, nous avons expliquée également les différentes étapes qui sont la collecte des données, création d'ontologie et analyse des informations. Dans l'analyse d'information, nous avons cité les différentes étapes comme Tokenisation, Sentence Splitter, et le Part Of Speech Tagger, analyseur Morphologique, OntoRootGazetteer et la machine Learning et des exemples pour plus de clarté.

Dans le chapitre suivant nous allons réaliser le modèle d'analyse sémantique des opinions proposées par des outils d'applications.

# **Chapitre 5 : Implémentation et Evaluation**

## 5.1 Introduction

Dans ce chapitre, nous allons mettre en œuvre les étapes de notre approche présentée précédemment. Nous commencerons par la conception d'un système d'extraction d'entités nommées basé sur une ontologie. Pour ce faire, nous utiliserons l'outil Protégé pour créer l'ontologie et nous choisirons la plateforme GATE comme environnement de développement de notre application. Ensuite, nous intégrerons les mêmes données extraites dans un processus de machine Learning en utilisant l'algorithme d'apprentissage par arbre de décision, implémenté en Python. Cette étape nous permettra de classifier et d'analyser les opinions exprimées sur la plateforme Facebook concernant le Bitcoin.

## 5.2 Protégé

Pour développer des ontologies, nous avons utilisé Protégé 5.6.3 [47]. Protégé est un éditeur d'ontologies destiné aux bases de connaissances, doté d'une interface utilisateur graphique facilitant la construction d'ontologies. Cet outil offre une flexibilité notable pour la méta-modélisation, permettant ainsi la création d'ontologies spécifiques à un domaine et la personnalisation des formulaires de saisie des données. Il est principalement utilisé dans les domaines de l'ingénierie des connaissances et de la modélisation conceptuelle [48].

La figure 5.1 présente la hiérarchie des classes et les graphiques associés dans Protégé.

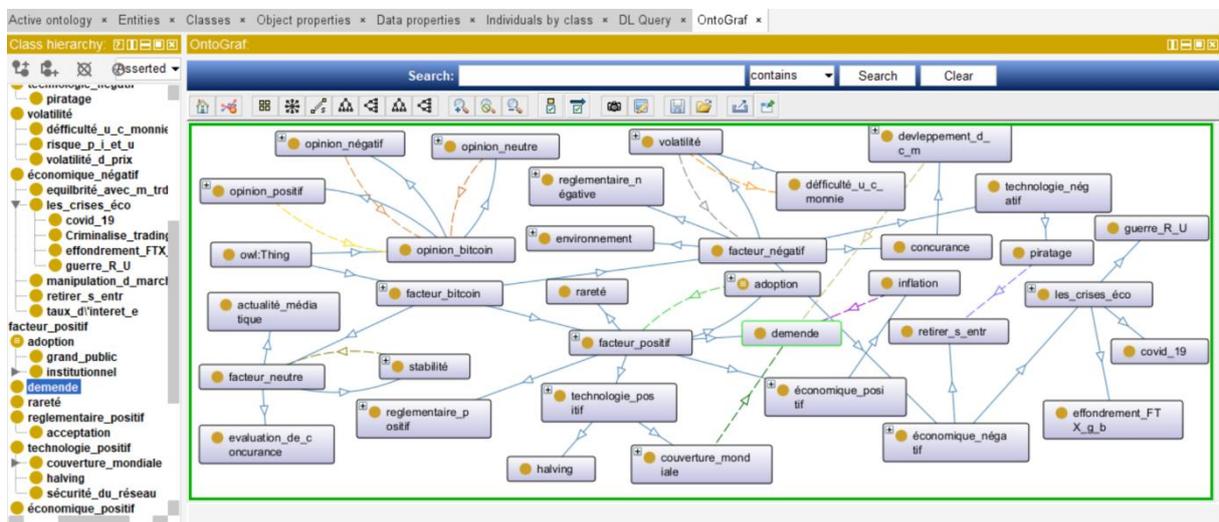


Figure 5.1: Protégé capture d'écran.

### 5.2.1 Création de l'ontologie

Comme nous l'avons déjà démontré lors de la création d'ontologies, le développement manuel d'ontologies se déroule selon les étapes suivantes :

#### **Étape 1 : Déterminer le domaine et la portée de l'ontologie**

La définition du domaine et de la portée de l'ontologie est la première étape du développement de l'ontologie. Afin d'évaluer le domaine et la portée de l'ontologie, il est essentiel de connaître le domaine couvert par l'ontologie, son objectif et le type de question auxquelles les informations contenues dans l'ontologie peuvent apporter des réponses, ainsi que l'utilisation de l'ontologie, notre domaine d'expertise est les facteurs et les opinions de Bitcoin .

#### **Étape 2 : Envisager de réutiliser les ontologies existantes**

Dans cette étape, il est nécessaire de vérifier l'existence d'une ontologie développée auparavant dans le domaine. Si une ontologie comme celle-ci existe, il est souvent plus aisé de la modifier pour répondre à ses propres besoins que de la créer. Toutefois, dans notre situation, nous avons négligé cette étape car nous n'avons pas découvert d'ontologie préexistante liée à notre domaine, les facteurs qui influençant la valeur du Bitcoin .

#### **Étape 3 : Énumérer les termes importants dans l'ontologie**

On peut voir dans cette étape une activité de remue-méninges, la liste des mots que nous souhaitons utiliser, la démonstration des termes de l'ontologie et des propriétés qu'ils peuvent avoir, l'utilisation des documents collectés pour l'acquisition des connaissances sur les facteurs de Bitcoin.

#### **Étape 4 : Définir les classes et la hiérarchie des classes**

Cette étape peut être perçue comme une activité de réflexion, où nous énumérons les mots que nous souhaitons utiliser afin de mettre en évidence les termes de l'ontologie et les caractéristiques qu'ils peuvent posséder.

La figure 5.2 représente les classes d'ontologie de niveau supérieur. La notion de chose est représentée comme une classe qui englobe toutes les entités. La classe (Bitcoin) joue le rôle de classe racine, tandis que les autres classes agissent comme ses sous-classes.

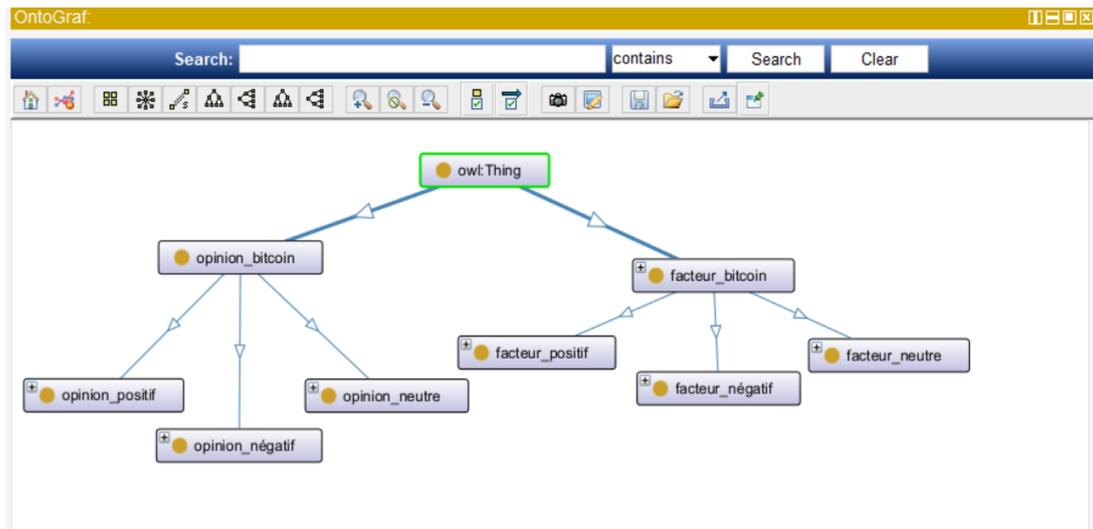


Figure 5.2 : Classes d'ontologie de haut niveau

### Étape 5 : Définir les propriétés des classes (slots).

Déterminer les propriétés d'objet (relations) entre les classes, il s'agit de relier les concepts de l'ontologie.

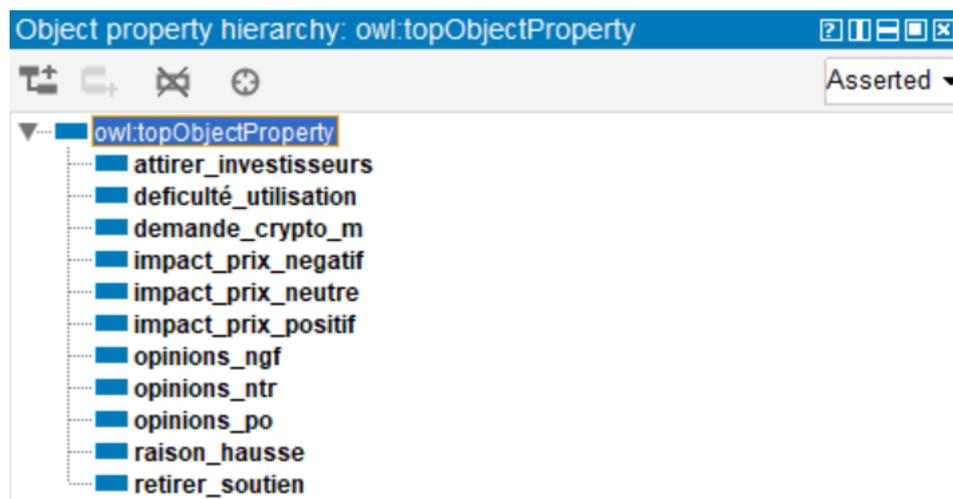


Figure 5.3 : Hiérarchie de propriété d'objet.

### Étape 6 : Définir les facettes des fentes

Les emplacements (parfois désignés sous le nom de rôles ou de propriétés) présentent divers aspects (parfois désignés sous le nom de restrictions de rôle) qui décrivent le type de valeur, les valeurs autorisées, le nombre de valeurs (cardinalité) et d'autres caractéristiques des valeurs que peut prendre l'emplacement. Par exemple, une catégorie est liée à une relation qui possède une valeur.

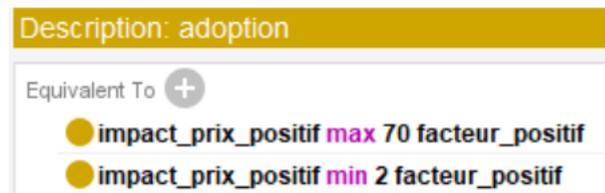


Figure 5.4 : Exemple de restrictions de rôle.

### Étape 7: Créer des instances

La création d'instances (individus) revêt une grande importance pour enrichir l'ontologie, car elle est directement liée aux classes et aux sous-classes.

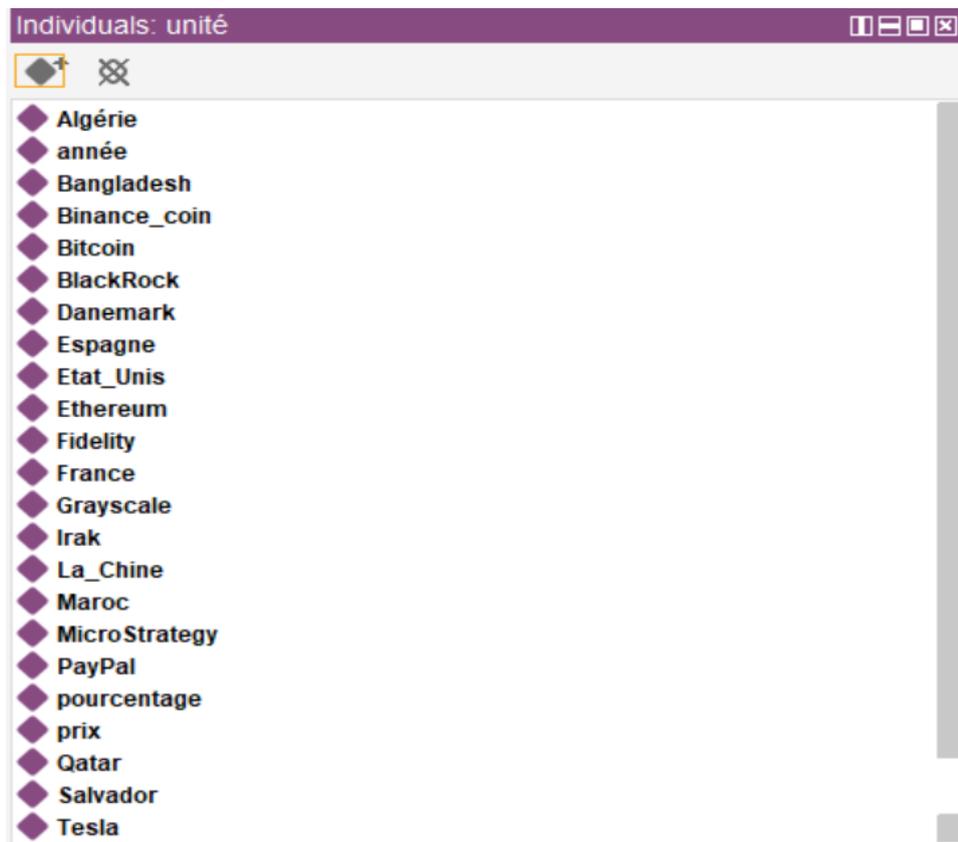


Figure 5.5 : Liste d'instances d'ontologie.

## 5.3 GATE

GATE, ou General Architecture for Text Engineering, est un logiciel open source créé par l'Université de Sheffield. Ce logiciel fournit une infrastructure pour le développement et le déploiement de composants logiciels spécialisés dans le traitement du langage humain. Reconnue comme un outil performant dans le domaine du traitement automatique du langage naturel (NLP), GATE propose de nombreuses fonctionnalités, notamment l'annotation manuelle et automatique, l'utilisation de diverses ressources lexicales (gazetteers), la récupération d'informations, et le traitement basé sur les ontologies.

GATE est composé de divers éléments qui peuvent être classés en trois types différents:

- **Ressources langage (LRs : LanguageResources)**

C'est une variété de données linguistiques comme des documents, des corpus, des lexiques ou des ontologies.

- **Ressources de traitement (Algorithmique) (PRs : Processing Resources)**

Il s'agit de programmes ou d'algorithmes qui effectueront une sorte de traitement sur le texte, tels que la tokenisation, le dictionnaire de recherche, l'analyse, etc. En général, les PR sont employés afin de traiter les informations fournies par les LR.

- **Ressources de visualisation (VRs : Visual Resources)**

Ces éléments visuels offrent la possibilité de présenter les résultats à l'intérieur de l'environnement de développement GATE et d'éditer d'autres types de ressources (Figure 5.6).

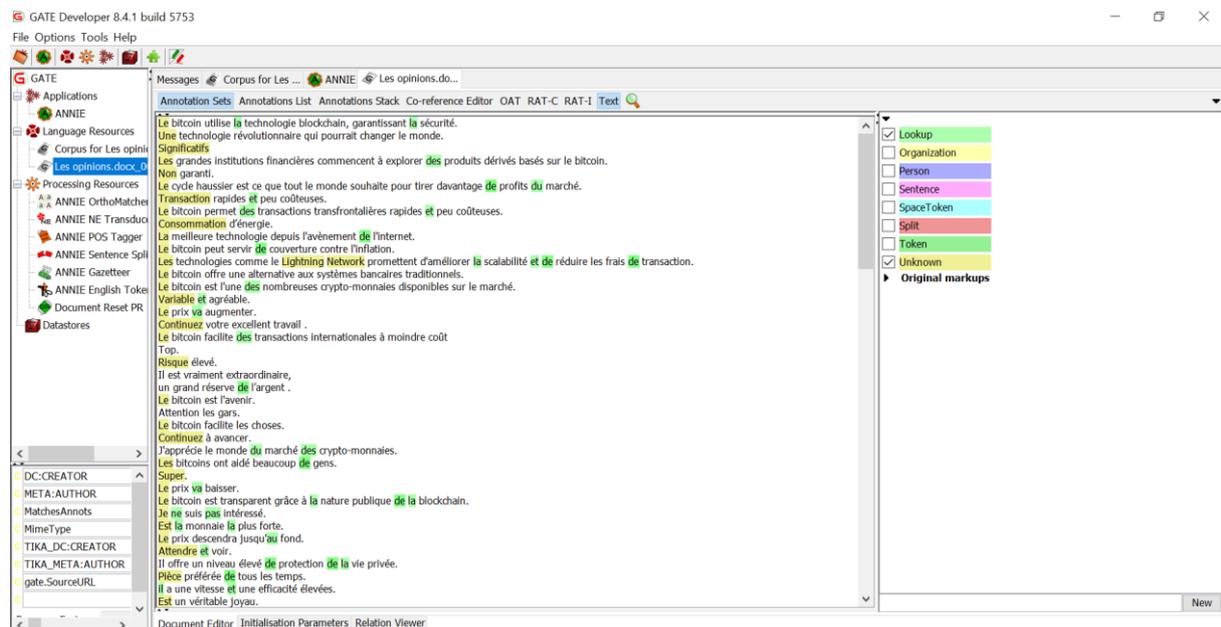


Figure 5.6: Exemple de l'interface dans GATE.

## 5.4 ANNIE

ANNIE (A Nearly-New Information Extraction System), est un élément de GATE, composé de divers modules tels qu'un analyseur lexical (Tokenizer), un Gazetteer (liste de mots-clés), un segmenter de phrases, un étiqueteur, un module d'extraction d'entités nommées et un module de détection de corrélation. ANNIE propose toutes les ressources de traitement requises pour analyser les informations présentes dans les textes (Extraction d'informations). Il propose également des outils pour traiter les phrases, détecter les entités et détecter les références entre les sections d'un texte. Il se compose d'un :

### 5.4.1 Tokeniser (TokeniserPR)

Ce module permet de séparer le texte en éléments simples (ponctuations, nombres, mots, etc) [44].

### 5.4.2 Gazetteer (Gazetteer PR)

Ce qui est une collection de listes. Le rôle du module de Gazetteer consiste à repérer les noms d'entités dans le texte en se basant sur les listes disponibles. On retrouve dans ces listes, par exemple, tous les noms de villes ou de pays. Chaque liste est composée d'un ensemble de noms, comme les noms de villes, d'organisations, les jours de la semaine, notamment [44].

### 5.4.3 Sentence Splitter

Il divise le texte qui lui est donné en entrée en un ensemble de phrases en fonction de la ponctuation correspondante. En effet, il s'agit d'extraire des informations phrase par phrase. Il n'est pas possible d'utiliser d'autres informations que celles de cette phrase lors du processus. Ce module consiste en une série de transducteurs à états finis qui divisent le texte en phrases distinctes. Il est nécessaire pour la section Tagger de la parole (PoS) [44].

### 5.4.4 POS Tagger

Qui est responsable de la correction grammaticale du texte. Le POS Tagger utilisé par GATE est une version modifiée du tagger de Brill. Il génère une étiquette pour chaque mot ou symbole du discours sous la forme d'une annotation [44].

### 5.4.5 Named-Entity transducer (NE transducer)

Il s'agit de la partie de l'algorithme qui va exploiter toutes les informations précédentes afin de localiser les entités identifiées [44].

## 5.5 Différentes méthodes appliquées

Les techniques d'extraction qui ont été détaillées précédemment sont utilisées dans l'environnement de traitement de texte GATE. La base de nos chaînes d'extraction de relations est utilisée pour extraire les entités opinions provenant d'ontologies.

Nous avons comparé trois méthodes pour extraire les opinions, à savoir :

### 5.5.1 Méthode par défaut

Dans cette application, les documents sont traités en utilisant l'application ANNIE par défaut, ce qui donne les résultats suivants :

L'exécution d'un document en utilisant l'application par défaut, le système ne détecte aucune entité nommée « Opinion Bitcoin » concernant une opinion sur Bitcoin. Le résultat est comme suit :

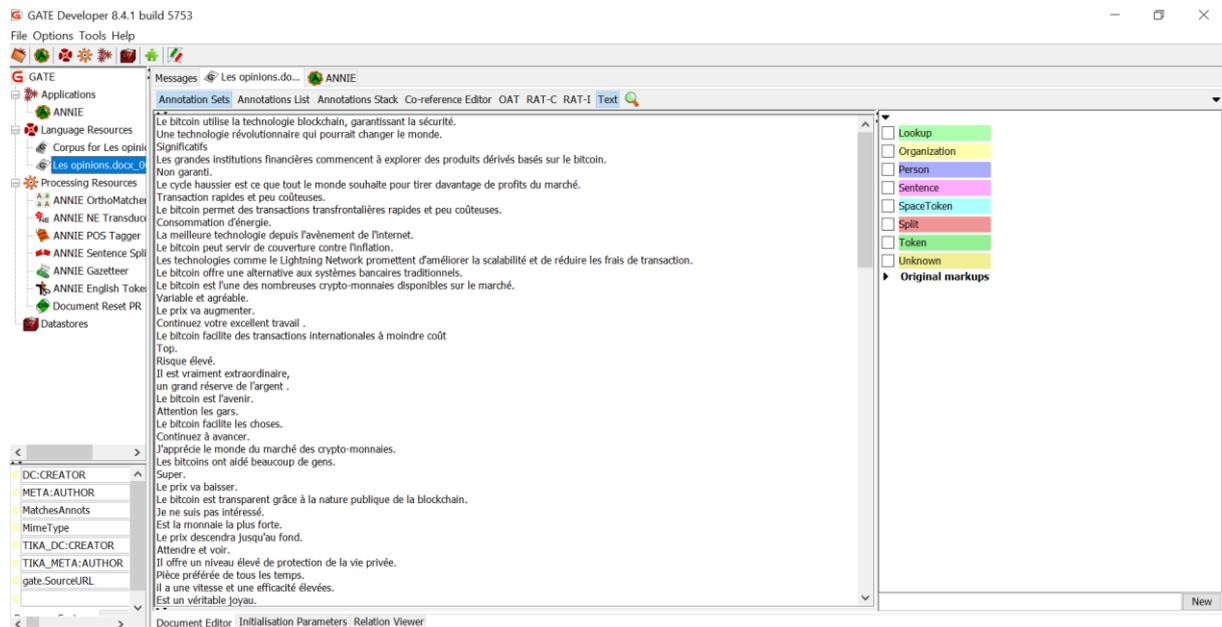


Figure 5.7 : Exécution Opinion Bitcoin par Default.

### 5.5.2 Méthode basé par l'ontologie

Dans cette méthode nous utilisons l'ontologie pour connaitre les différentes opinions sur Bitcoin dans le texte traité. Le système explorera notre ontologie de domaine et présentera les résultats de la manière suivante :

En exécutant le document en utilisant l'ontologie, le système détecte plus des opinions «Opinion Bitcoin», Le résultat était comme suit :

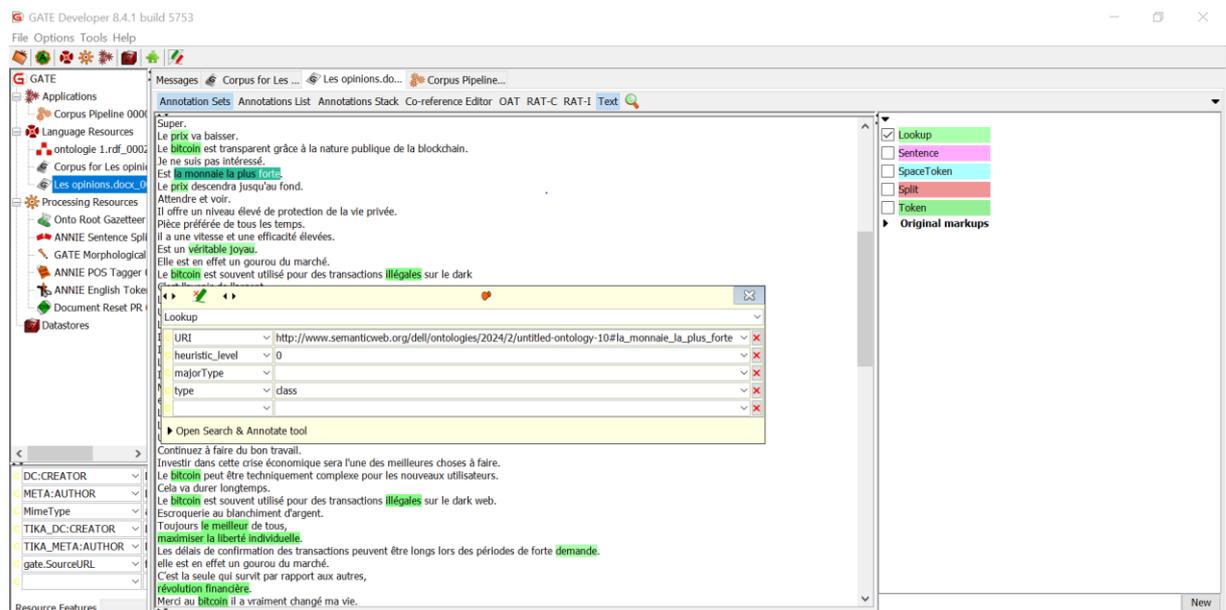
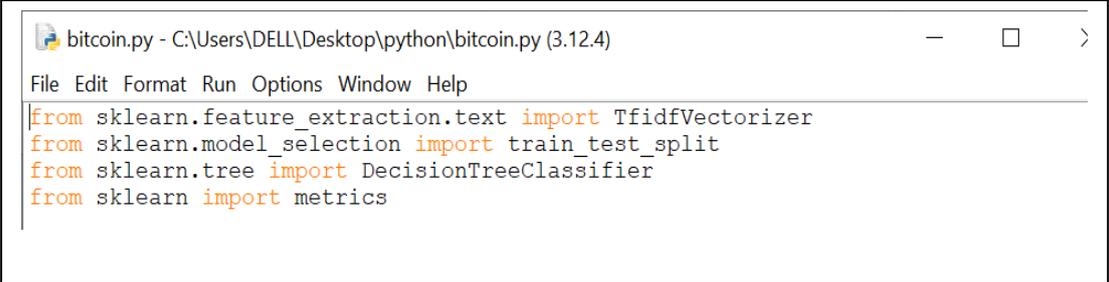


Figure 5.8: Exécution Opinion basé ontologie (lookup) (Opinion Bitcoin).

### 5.5.3. Méthode basé par la machine Learning

Dans cette méthode, nous avons utilisé la machine Learning pour traiter les mêmes opinions de Bitcoin en utilisant l'algorithme de l'arbre de décision, afin de les comparer avec la méthode basée sur l'ontologie. Nous avons utilisé le langage de programmation python, ce programme utilise plusieurs bibliothèques Python pour effectuer l'analyse de sentiments à l'aide d'un arbre de décision. La capture suivante représente (5.9) la bibliothèque (sklearn,..), les modules (feature\_extraction,..) et les classes (TfidfVectorizer,...) utilisées dans notre travail :



```
bitcoin.py - C:\Users\DELL\Desktop\python\bitcoin.py (3.12.4)
File Edit Format Run Options Window Help
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn import metrics
```

**Figure 5.9 :** Packages de Python.

#### 1) **sklearn.feature\_extraction.text.TfidfVectorizer**

Cette classe est utilisée pour transformer les textes en matrices de termes pondérés par TF-IDF (Term Frequency-Inverse Document Frequency).

#### 2) **sklearn.model\_selection.train\_test\_split**

Ce modèle est utilisé pour séparer les données en ensembles d'entraînement et de test, ce qui est essentiel pour évaluer la performance du modèle.

#### 3) **sklearn.tree.DecisionTreeClassifier**

Cette classe fournit l'implémentation de l'algorithme de l'arbre de décision pour la classification.

#### 4) **sklearn.metrics**

Cette classe contient des fonctions pour évaluer la performance du modèle, telles que `classification_report` pour obtenir un rapport de classification détaillé et `accuracy_score` pour calculer l'exactitude du modèle.

L'algorithme de l'arbre de décision à plusieurs avantages, nous en mentionnons quelques-uns :

- ✓ Simplicité d'implémentation.
- ✓ Interprétabilité élevée.
- ✓ Efficacité avec de grands ensembles de données.
- ✓ Sélection automatique des caractéristiques.

Dans notre travail, nous avons utilisé cet algorithme car il est particulièrement efficace pour l'extraction des opinions et surpassant d'autres algorithmes d'apprentissage automatique. Il est facile à comprendre et à visualiser, adapter aux données textuelles vectorisées, simple à implémenter et efficace pour capturer des relations complexes entre les caractéristiques et les sentiments. Nous avons exploité les paramètres suivants dans notre travail avec l'algorithme de l'arbre de décision :

- ✓ TfidfVectorizer pour la vectorisation des Textes.
- ✓ train\_test\_split pour la séparation des données.
- ✓ DecisionTreeClassifier pour l'entraînement du modèle.

La capture suivante (5.10) représente l'exécution de la machine Learning :

```
===== RESTART: C:/Users/DELL/Desktop/python/testcor.py =====
Number of text samples: 173
Number of labels: 165
      precision    recall  f1-score   support

-1         0.67     0.20     0.31         10
 0         0.60     0.83     0.70         18
 1         0.60     0.60     0.60          5

 accuracy          0.61         33
 macro avg         0.62         33
 weighted avg      0.62         33

Accuracy: 0.60606060606061
|
```

**Figure 5.10:** Exécution de la machine Learning.

## 5.6 Evaluation

Nous avons examiné un corpus comprenant 173 commentaires sur les opinions de Bitcoin issus des réseaux sociaux (Facebook) en utilisant la première méthode de l'ontologie et la deuxième méthode de la machine Learning.

Nous exposons différentes méthodes d'évaluation comme la précision, le rappel et la mesure F, qui sont couramment employées dans les évaluations en TALN ( Traitement Automatique des langages Naturel).

- La précision est le pourcentage des résultats corrects parmi les résultats obtenus.
- Le rappel est le pourcentage des résultats corrects parmi les résultats qu'on doit trouver.
- Pour le domaine de l'extraction des ENS, les taux de la précision et du rappel sont calculés selon les formules suivantes :

$$\text{Précision} = \frac{\text{Nombre d'ENS correctement reconnus}}{\text{Nombre d'ENS reconnues}}$$

$$\text{Rappel} = \frac{\text{Nombre d'ENS correctement reconnus}}{\text{Nombre d'ENS de corpus}}$$

La F-mesure est la combinaison de la précision et du rappel et leur pondération.

La formule de la F-mesure est comme suit [49] :

$$F - \text{mesure} = \frac{2(\text{précision} * \text{rappel})}{\text{rappel} + \text{précision}}$$

Pour évaluer les performances de deux méthodes d'extraction des opinions (une basée sur l'ontologie et l'autre sur la machine Learning), nous examinons les mesures de précision, rappel, et F-mesure pour les catégories d'opinions positives, négatives et neutres. Voici une analyse détaillée des résultats obtenus.

**Tableau 5.1** : Evaluation Opinion Positive

Catégorie	Précision	Rappel	F-mesure
Méthode d'ontologie	0,56	0,74	0,63
Méthode Machine Learning	0,60	0,60	0,60

#### Analyse :

- **Précision** : La méthode de machine Learning affiche une précision légèrement supérieure (0,60) par rapport à la méthode basée sur l'ontologie (0,56). Cela signifie que la machine Learning est légèrement meilleure pour identifier correctement les opinions positives.
- **Rappel** : La méthode d'ontologie a un rappel plus élevé (0,74) comparé à la machine Learning (0,60). Cela indique que l'ontologie est plus efficace pour retrouver la majorité des opinions positives présentes dans les données.

- **F-mesure:** L'ontologie a une F-mesure de 0,63, légèrement supérieure à celle de la machine Learning (0,60), suggérant une meilleure balance entre précision et rappel pour cette catégorie.

**Tableau 5.2 :** Evaluation Opinion Négative

Catégorie	Précision	Rappel	F-mesure
Méthode d'ontologie	0,62	0,85	0,71
Méthode Machine Learning	0,67	0,20	0,31

**Analyse :**

- **Précision:** La méthode de machine Learning montre une précision légèrement supérieure (0,67) par rapport à la méthode d'ontologie (0,62).
- **Rappel:** La méthode d'ontologie surpasse largement la machine Learning avec un rappel de 0,85 contre 0,20. Cela signifie que la méthode d'ontologie est bien meilleure pour identifier toutes les opinions négatives dans les données.
- **F-mesure:** La F-mesure de l'ontologie (0,71) est nettement supérieure à celle de la machine Learning (0,31), indiquant une meilleure performance globale pour cette catégorie.

**Tableau 5.3 :** Evaluation Opinion Neutre

Catégorie	Précision	Rappel	F-mesure
Méthode d'ontologie	0,57	0,80	0,66
Méthode Machine Learning	0,60	0,83	0,70

**Analyse :**

- **Précision:** La méthode de machine Learning affiche une précision légèrement supérieure (0,60) par rapport à l'ontologie (0,57).
- **Rappel :** Le rappel de la machine Learning (0,83) est également légèrement supérieur à celui de l'ontologie (0,80).
- **F-mesure :** La F-mesure de la machine Learning (0,70) est légèrement supérieure à celle de l'ontologie (0,66), suggérant que la machine Learning est légèrement meilleur

pour capturer et équilibrer les opinions neutres.

D'après les résultats obtenus, nous avons fait une comparaison entre les deux méthodes :

### 1) Méthode d'Ontologie

- **Points Forts :** La méthode basée sur l'ontologie est particulièrement efficace pour les opinions négatives, avec un rappel et une F-mesure élevés. Cela peut être dû à la capacité de l'ontologie à capturer des relations contextuelles complexes et spécifiques.
- **Points Faibles:** La précision est généralement inférieure à celle de la méthode de machine Learning pour toutes les catégories, ce qui indique une tendance à classer incorrectement certaines opinions.

### 2) Méthode de Machine Learning

- **Points Forts :** Cette méthode montre une meilleure précision pour les opinions positives et neutres. Pour les opinions neutres, elle offre également un rappel et une F-mesure légèrement meilleurs.
- **Points Faibles:** La méthode de machine Learning a une performance très faible pour les opinions négatives, avec un rappel particulièrement bas, ce qui indique des difficultés à identifier les opinions négatives de manière exhaustive.

## 5.7 Conclusion

Ce chapitre détaille la réalisation de notre approche proposée, en incluant une vue d'ensemble des langages et des outils utilisés pour développer notre application. Nous avons employé deux approches la première basée sur l'ontologie et le deuxième est basées sur la machine Learning pour extraire les opinions de Bitcoin. Nous avons cité les résultats avec les deux méthodes et nous avons fait une comparaison avec les deux méthodes.

# Conclusion générale

---

## Conclusion générale

Notre travail a pour objectif de faire l'analyse sémantique des opinions dans le domaine économique. Nous avons initié notre exploration en penchant sur les thèmes de recherche en relation avec l'analyse sémantique des opinions, les ontologies, la machine Learning et plus spécifiquement sur les thèmes qui abordent l'analyse sémantiques des opinions dans le domaine économique. Le document est devisé en cinq chapitres, chacun abordant devers aspects du sujet.

Le premier chapitre commence par une introduction sur l'analyse des sentiments, définissant ce concept. Par la suite nous avons présenté les domaines d'applications, ses niveaux et les difficultés liées à l'analyse des opinions. Le chapitre se termine par une exploration des différentes approches de l'analyse des sentiments.

Le deuxième chapitre se focalise sur les ontologies et l'analyse sémantique des opinions. Nous avons donné une introduction aux ontologies et mis en évidence les classifications d'ontologies, puis nous avons présenté ses avantages et son utilisation.

Le troisième chapitre est consacré à l'analyse sémantique des opinions dans le domaine économique, en particulier concernant la monnaie numérique Bitcoin. Nous avons commencé par définir le Bitcoin, puis nous avons expliqué son fonctionnement et ses caractéristiques, ainsi que ses avantages et ses inconvénients. Le chapitre se termine par une revue des travaux liés à cette thématique.

Le quatrième chapitre présente un modèle de conception pour un système de détection sémantique des opinions. Nous avons commencé par proposer notre approche. Ensuite, nous avons détaillé l'architecture du système, ainsi que les étapes telles que la collecte des données, le prétraitement, la création d'une ontologie de domaine et l'analyse des informations. Nous avons expliqué la méthode de machine Learning, en commençant par son utilisation sur la plateforme GATE. Nous avons ensuite présenté notre proposition et défini l'algorithme de l'arbre de décision utilisé pour la comparaison avec la méthode basée sur l'ontologie.

Le dernier chapitre traite la mise en œuvre de notre solution proposée, mettant en lumière l'utilisation de l'outil Protégé pour construire l'ontologie et de l'outil GATE pour analyser les textes. Diverses approches sont adoptées, notamment la méthode par défaut, la

## **Conclusion générale**

---

méthode basée sur l'ontologie et la méthode basée sur la machine Learning qui a utilisé l'algorithme de l'arbre de décision. En outre, une évaluation est effectuée pour déterminer les performances de notre solution.

En conclusion, cette étude a abouti à la proposition d'un modèle pour l'analyse sémantique des opinions dans le domaine économique. Nous avons effectué une analyse approfondie des concepts fondamentaux, des méthodes existantes et des recherches connexes dans ce domaine.

Nous avons mis en œuvre notre solution en utilisant les outils Protégé, GATE et Python puis évalué les résultats obtenus. Cette contribution marque une avancée significative dans la recherche sur l'analyse des opinions dans le domaine économique.

## Bibliographies

- [1] Godsay, M. (2015). The process of sentiment analysis: a study. *International Journal of Computer Applications*, 126(7), 26-30.
- [2] Pak, A. (2012). *Automatic, adaptive, and applicative sentiment analysis* (Doctoral dissertation, Université Paris Sud-Paris XI).
- [3] Larousse, P. (1869). *Grand dictionnaire universel du XIXe siècle: français, historique, géographique, mythologique, bibliographique, littéraire, artistique, scientifique, etc...* (Vol. 4). Larousse et Boyer
- [4] Indurkha, N., & Damerou, F. J. (2010). *Handbook of natural language processing*. Chapman and Hall/CRC.
- [5] Sarkar, D. (2019). *Text analytics with Python: a practitioner's guide to natural language processing* (pp. 1-674). Bangalore: Apress.
- [6] Yuri, M. N., & Rosli, M. M. (2022). TelcoSentiment: Sentiment Analysis on Mobile Telecommunication Services. *Journal of Positive School Psychology*, 6(3), 8470-8481.
- [7] Duwairi, R. M., Marji, R., Sha'ban, N., & Rushaidat, S. (2014, April). Sentiment analysis in arabic tweets. In *2014 5th international conference on information and communication systems (ICICS)* (pp. 1-6). IEEE.
- [8] Poirier, D., Fessant, F., Bothorel, C., De Neef, E. G., & Boullé, M. (2009). Approches statistique et linguistique pour la classification de textes d'opinion portant sur les films. *Revue des Nouvelles Technologies de l'Information*, Pages-147.
- [9] <https://fastercapital.com/fr/contenu/Analyse-des-sentiments---comprendre-l-analyse-des-sentiments-dans-la-decouverte-des-prix.html#Outils-et-technologies-d-analyse-des-sentiments>
- [10] Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.
- [11] Kolkur, S., Dantal, G., & Mahe, R. (2015). Study of different levels for sentiment analysis. *International Journal of Current Engineering and Technology*, 5(2), 768-770.
- [12] [https://www.researchgate.net/publication/350993832\\_Analyse\\_de\\_sentiments\\_dans\\_les\\_textes\\_economiques\\_un\\_exemple\\_d\\_application\\_chez\\_ReportLinker](https://www.researchgate.net/publication/350993832_Analyse_de_sentiments_dans_les_textes_economiques_un_exemple_d_application_chez_ReportLinker)
- [13] Mullen, T., & Collier, N. (2004, July). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 412-418).
- [14] Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*.
- [15] <https://www.ibm.com/fr-fr/topics/semi-supervised-learning>
- [16] [https://fr.wikipedia.org/wiki/Apprentissage\\_non\\_supervis%C3%A9](https://fr.wikipedia.org/wiki/Apprentissage_non_supervis%C3%A9)

- [17] Hadji, M. (2019). Analyse des sentiments: Généralités. <https://di.univ-blida.dz/jspui/bitstream/123456789/11371/1/Mosteghanemi%20Samira%20%20et%20%20Fe%20roukhi%20Afaf.pdf>
- [18] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.
- [19] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.
- [20] <https://dynamics.microsoft.com/fr-ch/ai/customer-insights/what-is-sentiment-analysis/>
- [21] <https://www.wizville.com/ressources/lexique/analyse-semantique-avis-clients/>
- [22] <https://cordis.europa.eu/article/id/84136-ontology-approach-for-the-semantic-web/fr>
- [23] Fatima lamrache et Kheloudja Oukfif , « Analyse et classification d'opinionsur les produits ENIEM », Université Mouloud Mammeri de Tizi-Ouzou (2016).
- [24] Natalya F. Noy et Deborah L. McGuinness, développement d'une ontologie 101 : « Guide pour la création de votre première ontologie ». Stanford Knowledge systems Laboratory Technical Report mars (2011).
- [25] Laublet, P., Aussenac-Gilles, N., Camps, V., Glize, P., Hernandez, N., Maurel, H., ... & Tissaoui, A. (2009). DYNAMIC Ontology for information retrieval.
- [26] Aussenac-Gilles, N., & Mothe, J. (2004, April). Ontologies as Background Knowledge to Explore Document Collections. In *RIAO* (Vol. 4, pp. 129-142).
- [27] Ferroudja, B., Nawel, H., & Faiza, R. (2013). *Modèle de langue mixte combinant entités nommées et mots simples* (Doctoral dissertation, Université Mouloud Mammeri).
- [28] GHANINE Mohammed & MEDJBOUR Aghiles , « Construction d'une ontologie de domaine médical » , Université Mouloud Mammeri de Tizi-Ouzou (2015).
- [29] Psyché, V., Mendes, O., & Bourdeau, J. (2003). Apport de l'ingénierie ontologique aux environnements de formation à distance. *STICEF (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation)*, 10, 89-126.
- [30] Kiong, Y. C., Palaniappan, S., & Yahaya, N. A. (2011, July). Health ontology system. In *2011 7th International Conference on Information Technology in Asia* (pp. 1-4). IEEE.
- [31] Van Heijst, G., Schreiber, A. T., & Wielinga, B. J. (1997). Using explicit ontologies in KBS development. *International journal of human-computer studies*, 46(2-3), 183-292.
- [32] Mendes, O. (2003). Apport de l'ingénierie ontologique aux environnements de formation à distance.
- [33] Yeretjian, A., Jeanneau, C., Stachtchenko, A., & Balva, C. (2016). La blockchain

décryptée-Les clés d'une révolution.

[34]<https://www.xtb.com/fr/formation/Bitcoin>

[35]<https://www.cointribune.com/btc-comment-ca-marche/>

[36] <https://coinacademy.fr/Bitcoin-btc-fondamental/>

[37] [https://icesi.fr/11\\_Le-Bitcoin---Fonctionnement-et-succes.html](https://icesi.fr/11_Le-Bitcoin---Fonctionnement-et-succes.html)

[38] <https://www.presse-citron.net/quest-ce-que-le-Bitcoin/>

[39] <https://byjus.com/current-affairs/cryptocurrency/>

[40]<https://cs.stanford.edu/people/eroberts/cs181/projects/201011/DigitalCurrencies/disadvantages/index.html>

[41] <https://www.smob.fr/blog/accepter-les-paiements-en-Bitcoin-dans-votre-boutique-les-avantages-et-les-inconvenients/>

[42] Byc , K ., & Ilinca , S.C. (2021) .The Relationship Between Social Media Sentiment and Bitcoin Price Volatility (pp. 1-100).

[43]Bhatt, S., Ghazanfar, M., & Amirhosseini, M. (2023). Machine learning based cryptocurrency price prediction using historical data and social media sentiment. *Computer Science & Information Technology (CS & IT)*, 13(10), 1-11.

[44] <https://gate.ac.uk/sale/tao/split.html>

[45] <https://cedric.cnam.fr/vertigo/cours/ml2/tp Arbres Decision.html>

[46] <https://www.sciencedirect.com/topics/computer-science/decision-tree-classifier>

[47] <https://protege.stanford.edu/software.php>

[48] Kapoor, B., & Sharma, S. (2010). A comparative study ontology building tools for semantic web applications. *International journal of Web & Semantic Technology (IJWesT)*, 1(3), 1-13.

[49] Hengchen, S., Van Hooland, S., Verborgh, R., & De Wilde, M. (2015). L'extraction d'entités nommées: une opportunité pour le secteur culturel?. *I2D-Information, données & documents*, 52(2), 70-79.