

PFE

PROJET DE FIN D'ETUDES

Thème :

Construction d'un corpus en langue française
(Dictionnaire NLP) pour l'analyse de sentiments des textes
relevant des décisions de la politique monétaire de BANK
AL-MAGRIB

Entreprise

Bank Al-Maghrib

Filière

Génie logiciel

Etudiant

Imane Khayati

Encadrant

Dr. Oussama Smimite

Superviseur

M. Ahl Mbarek Salah

Dédicaces

Dans ce moment précieux, je voudrais prendre un instant pour exprimer mes dédicaces les plus sincères à celles qui ont illuminé mon chemin de leur présence bienveillante et de leur amour inconditionnel.

À mes chers parents, dont le soutien indéfectible a été ma force et ma boussole dans les moments de doute et de difficulté. Vos encouragements et vos sacrifices ont été le moteur de ma réussite, et pour cela, je vous serai éternellement reconnaissante.

À mes amies, compagnes de joie et de peine, qui ont partagé avec moi les hauts et les bas de cette aventure. Leur présence à mes côtés a rendu chaque étape de ce parcours plus lumineuse et plus significative.

À mes mentoras et enseignantes, dont la sagesse et les conseils éclairés ont nourri ma croissance personnelle et professionnelle. Leur guidance m'a permis de me surpasser et de réaliser mon plein potentiel.

Et enfin, à toutes les âmes inspirantes que j'ai rencontrées en chemin, qui ont enrichi ma vie de leur bienveillance, de leur gentillesse et de leur amour. Leurs sourires et leurs encouragements ont été les étoiles qui ont illuminé mes nuits les plus sombres.

À vous toutes, je dédie ces mots empreints de gratitude et d'amour. Que nos liens perdurent et que nos chemins se croisent à nouveau, porteurs de nouvelles aventures et de moments précieux à partager.

Remerciements

À cette occasion, je tiens sincèrement à exprimer ma reconnaissance envers toutes les personnes qui ont contribué au succès de ce travail, en particulier mes parents, qui ont veillé à ce que je dispose des meilleures conditions de travail, me soutenant de manière inconditionnelle tant sur le plan physique que mental. Je souhaite également exprimer ma profonde gratitude envers mes superviseurs et mentors, M. Ahl Mbarek Salah et M. Youssef Karim el Alaoui, qui m'ont offert leur confiance pour cette mission.

Je suis admirative et reconnaissante envers toute l'équipe et les personnels de la direction des statistiques et la gestion des données du Bank Al Maghrib, notamment Mlle. Sana Derqaoui et Mlle. Hiba Badiha avec lesquelles j'ai eu le privilège de collaborer et d'évoluer. Je leur souhaite à tous le plus grand des succès dans leurs projets futurs.

Par ailleurs, je tiens à exprimer ma gratitude envers l'École Polytechnique et son personnel pour avoir facilité notre apprentissage et nous avoir donné l'opportunité de nous développer. Mes remerciements vont également à nos professeures, sources d'inspiration et de sagesse. Je suis reconnaissante envers mes amies proches dont le soutien a rendu cette année inoubliable et enrichissante en souvenirs mémorables.

Enfin, je tiens à reconnaître mes propres efforts, ma bienveillance et ma passion dédiés à la réalisation de ce travail. Mon engagement et ma persévérance tout au long de ce processus ont été déterminants pour concrétiser ce projet. J'espère pouvoir continuer à accomplir de grandes réalisations à l'avenir.

Table des matières

Dédicaces	I
Remerciements	II
Table des matières	III
Liste des figures	VI
Liste des tableaux	VI
Abréviations / termes à définir	VII
Résumé	VIII
Abstract	IX
Introduction générale.....	1
Chapitre I : Contexte du PFE	3
1. Introduction	4
2. Organisation d'accueil.....	4
2.1. Présentation de l'organisation	4
2.2. Fonctions et missions	4
2.3. Organigramme de l'organisation.....	5
3. Cadre de la politique monétaire.....	5
3.1. Structure du système financier marocain	5
3.2. La politique monétaire.....	8
3.3. Les instruments de la politique monétaire.....	9
3.4. Décisions récentes de politique monétaire et leurs impacts	10
4. Cadre général du stage	11
5. Thématique.....	12
5.1. Présentation du sujet.....	12
5.2. Travail demandé.....	12
5.3. Délimitations	13
5.4. Objectifs à atteindre.....	13
5.5. Planification du projet	14
6. Conclusion.....	15
Chapitre II : Problématique et analyse	16
1. Introduction	17
2. Problématique et positionnement	17
2.1. Contexte général.....	17
2.2. Problématique.....	18
2.3. Analyse de l'existant	18
2.4. Critique de l'existant	20
3. Solutions proposées.....	22
4. Spécification des besoins.....	23
4.1. Besoins fonctionnels.....	23

4.2. Besoins non-fonctionnels	23
5. Méthodologie de travail.....	23
6. Conclusion.....	24
Chapitre III : Recherche bibliographique	25
1. Introduction	26
2. Contexte du l’art.....	26
2.1. Traitement automatique des langages naturels	26
2.1.1. Définition du NLP	26
2.1.2. Utilité du NLP	26
2.1.3. Le workflow du NLP.....	27
2.2. L’analyse des sentiments.....	28
2.3. Techniques de NLP	29
2.3.1. Tokenisation	29
2.3.2. Suppression du bruit.....	29
2.3.3. Les stopwords.....	30
2.3.4. Lemmatisation	30
2.3.5. Stemming.....	31
2.3.6. Reconnaissance d'entités nommées (NER)	31
2.3.7. Feature Engineering	32
2.4. Les modèles de machine learning.....	33
2.4.1. Naive bayes	33
2.4.2. SVM (Support Vector Machine)	34
3. Outils et logiciels.....	36
3.1. Excel.....	36
3.2. Visual Studio Code (VSC)	36
3.3. PyCharm.....	36
3.4. Lucidchart.....	36
3.5. Google Colab.....	36
4. Frameworks et langages de programmations	37
4.1. Node.js.....	37
4.2. Next.js	37
4.3. Python.....	37
4.4. Tailwind CSS	37
4.5. Bibliothèques Python	37
5. Conclusion.....	39
Chapitre IV : Ingénierie et analyse conceptuelle.....	40
1. Introduction	41
2. Conception	41
2.1. Diagramme en cas d’utilisation.....	41
2.2. Diagramme d’activité.....	42
3. Collecte des données	43

4.	Prétraitement des données	44
5.	Visualisation des données	45
6.	Résultat de l'entraînement et tests du modèle	46
6.1.	Définition des mesures de performances :.....	46
6.2.	Les résultats obtenus	46
6.2.1.	Naive Bayes.....	46
6.2.2.	SVM	47
6.2.3.	Confusion matrix	48
6.3.	Pourquoi Naive Bayes ?	49
7.	Conclusion.....	49
Chapitre V : Réalisation		51
1.	Introduction	52
2.	Les interfaces graphiques	52
2.1.	Formulaire d'Entrée	52
2.2.	Affichage des Résultats	52
3.	Conclusion.....	54
Conclusion générale		55
Bibliographie		57
Annexes.....		58

Liste des figures

Figure 1: Organigramme de Bkam.....	5
Figure 2: l'effet de la politique monétaire	9
Figure 3: Diagramme de Gantt.....	15
Figure 4: Kaggle- French corpus for sentiment analysis	19
Figure 5: Tatoeba dictionary	19
Figure 6: sketch engine - corpus maker.....	20
Figure 7: Natural language processing	26
Figure 8: workflow du NLP	27
Figure 9: diagramme en cas d'utilisation.....	41
Figure 10: diagramme d'activité du processus du projet	43
Figure 11: observation des articles par type de classification	44
Figure 12: processus du prétraitement des données	44
Figure 13: word-cloud des 2 articles du Corpus.....	46
Figure 14: matrice de Confusion Naive Bayes	48
Figure 15: matrice de confusion SVM	48
Figure 16: formulaire d'entrée	52
Figure 17: composants de la table de bord	53

Liste des tableaux

Tableau 1:structure du système financier marocain	6
Tableau 2: étude comparative des méthodes existantes	21
Tableau 3: l'importance d'un corpus dans le contexte de ce projet.....	22
Tableau 4: les étapes de l'analyse des sentiments.....	28
Tableau 5: bibliothèques de python.....	37
Tableau 6: metrics du Naive Bayes	46
Tableau 7: Metrics du SVM	47
Tableau 8: Etude comparative des modèles de ML.....	49

Abréviations / termes à définir

Terme	Définition
Corpus	Un ensemble de documents (des textes) regroupés dans une thématique précise.
Document	La notion de document fait référence à un texte appartenant au corpus, mais indépendant des autres textes. Il peut être constitué d'une ou plusieurs phrases, un ou plusieurs paragraphes.
Token	Désigne généralement un mot et/ou un élément de ponctuation. La phrase "Hello World !" comprend donc 3 tokens.
BAM / Bkam	Bank al-Maghrib
NLP	Natural Language Processing (Traitement automatique du langage naturel)
SVM	Support Vector Machine (Machine à vecteurs de support)
TF – IDF	Term Frequency – Inverse Document Frequency (Fréquence-Inverse Document Fréquence)
NER	Named Entity Recognition (Reconnaissance d'entités nommées)
NLTK	Natural Language Toolkit (Boîte à outils de traitement du langage naturel)

Résumé

Ce rapport présente en détail le travail effectué dans le cadre du projet de fin d'étude, mené sous la supervision de la Direction des Statistiques et de la Gestion des Données au sein de Bank Al-Maghrib, Siège Rabat.

Sous la supervision attentive de M. Ahl Mbarek Salah et M. Youssef Karim El Alaoui, cet ouvrage explore minutieusement les mécanismes de construction d'un corpus en langue française, mettant en lumière les enjeux fondamentaux liés à l'analyse des sentiments dans le contexte spécifique de la politique monétaire au Maroc.

L'objectif principal de cette étude était d'approfondir la compréhension des implications économiques des décisions politiques en matière monétaire, dans un contexte national marqué par des défis et des dynamiques économiques complexes. Ces décisions, influant directement sur des aspects cruciaux tels que l'inflation, la liquidité, le chômage et l'investissement, ont été analysées dans le contexte de l'évolution récente de la situation économique du Maroc.

Pour ce faire, des méthodes avancées d'analyse de texte et de Machine Learning ont été appliquées sur un ensemble exhaustif d'articles de presse pertinents. Ces méthodes, notamment le traitement automatique du langage naturel (NLP), ont permis d'extraire des informations pertinentes des textes, mettant en lumière les nuances et les tendances émergentes.

Dans ce cadre, ce rapport offre une analyse approfondie des tendances et des perceptions liées à la politique monétaire au Maroc, s'inscrivant dans un contexte national et international en perpétuelle évolution. Il ouvre également la voie à des pistes de recherche prometteuses pour le développement futur de l'économie et de l'analyse de texte dans le paysage économique marocain.

Mots clés : Bank Al-Maghrib, politique monétaire, corpus, analyse des sentiments, NLP

Abstract

This present document presents a detailed overview of the work conducted as part of the final year project, carried out under the supervision of the Statistics and Data Management Administration at Bank Al-Maghrib, Rabat Headquarters.

Under the careful guidance of Mr. Ahl Mbarek Salah et Mr. Youssef Karim El Alaoui, this document thoroughly explores the mechanisms involved in constructing a corpus in the French language, shedding light on the fundamental issues associated with sentiment analysis in the specific context of monetary policy in Morocco.

The primary objective of this study was to deepen the understanding of the economic implications of monetary policy decisions, within the national context marked by complex economic challenges and dynamics. These decisions, directly impacting crucial aspects such as inflation, liquidity, unemployment, and investment, were analyzed against the backdrop of the recent evolution of the economic situation in Morocco.

To achieve this goal, advanced methods of text analysis, sentiment analysis, and machine learning were applied to a comprehensive set of relevant press articles. These methods, including natural language processing (NLP), extracted valuable insights from the texts, highlighting nuances and emerging trends.

In conclusion, this report provides an in-depth analysis of trends and perceptions related to monetary policy in Morocco, within a constantly evolving national and international context. It also paves the way for promising research avenues for the future development of economics and text analysis within the Moroccan economic landscape.

Keywords: Bank Al-Maghrib, monetary policy, corpus, sentiment analysis, NLP

Introduction générale

Dans un monde de plus en plus interconnecté, où l'information circule sans entraves à travers les frontières, l'importance de comprendre et d'analyser les phénomènes économiques n'a jamais été aussi cruciale. Alors que les nations naviguent à travers des paysages économiques complexes caractérisés par la volatilité, l'incertitude et les avancées technologiques rapides, la nécessité de cadres robustes et d'outils analytiques pour décrypter les tendances économiques et les implications politiques devient impérative.

Dans ce contexte, le Royaume du Maroc se présente comme une étude de cas convaincante. Situé au carrefour de l'Afrique et de l'Europe, le Maroc bénéficie d'un paysage économique diversifié façonné par une myriade de facteurs, notamment sa situation géographique, son héritage historique et ses dynamiques socio-politiques. Au fil des ans, l'économie marocaine a connu des transformations significatives, portées par des programmes de développement ambitieux, des réformes stratégiques et un accent croissant sur l'innovation et la durabilité.

Au cœur du cadre économique marocain se trouve sa politique monétaire, orchestrée par la prestigieuse institution de Bank Al-Maghrib. Chargée de maintenir la stabilité des prix, de promouvoir la croissance économique et de sauvegarder l'intégrité du système financier, Bank Al-Maghrib joue un rôle central dans l'orientation de la trajectoire économique du pays. Dans le contexte de l'évolution des tendances mondiales et des dynamiques régionales, la formulation et la mise en œuvre de la politique monétaire au Maroc représentent un effort multifacette, mêlant théorie économique et pragmatisme du monde réel. En tant qu'institution centrale de la politique monétaire et financière du Maroc, Bank Al-Maghrib joue un rôle crucial dans la formulation et la mise en œuvre des politiques économiques du pays. Nous examinons le mandat, la structure organisationnelle et les mécanismes opérationnels de la banque centrale, ainsi que son influence sur les marchés financiers, la stabilité économique et la croissance du pays.

Nous analysons également la politique monétaire récente au Maroc et les développements économiques pertinents qui ont façonné cette politique. Nous abordons les décisions récentes de Bank Al-Maghrib concernant les taux d'intérêt, la liquidité du marché et d'autres mesures politiques visant à stabiliser l'économie et à promouvoir la croissance, ainsi que les défis spécifiques auxquels est confrontée la politique monétaire marocaine, tels que la maîtrise de l'inflation, la gestion de la dette publique et les fluctuations des taux de change. En examinant ces questions, nous cherchons à comprendre l'impact de la politique monétaire sur l'économie marocaine dans un contexte mondial en mutation, ainsi que les perspectives pour l'avenir de la politique monétaire et financière du pays.

La présente étude s'efforce d'explorer en profondeur les subtilités du paysage de la politique monétaire

au Maroc, avec un accent particulier sur l'analyse des données textuelles relatives aux décisions de politique monétaire. À travers une lentille interdisciplinaire qui fusionne économie, science des données et analyse linguistique, cette recherche vise à découvrir des perspectives nuancées sur les sentiments, les perceptions et les implications inscrites dans le discours entourant la politique monétaire au Maroc.

En exploitant des techniques avancées de traitement du langage naturel (NLP), d'analyse des sentiments et d'apprentissage automatique, cette étude cherche à extraire des renseignements pertinents à partir de sources textuelles telles que des articles de presse, des déclarations de politique et des rapports économiques. À travers une analyse empirique rigoureuse et une interprétation méticuleuse, les conclusions de cette recherche visent à enrichir notre compréhension des moteurs, des défis et des résultats associés à la politique monétaire dans le contexte marocain.

De plus, cette étude aspire à contribuer au discours plus large sur l'économie et l'analyse de texte, offrant des perspectives méthodologiques, des résultats empiriques et des recommandations politiques qui résonnent au-delà des frontières du Maroc. En comblant les lacunes entre les mondes universitaires, la formulation des politiques et la pratique industrielle, cette recherche vise à favoriser le dialogue, à stimuler l'innovation et à inspirer les futures générations de chercheurs, de praticiens et de décideurs.

Dans les sections suivantes de ce rapport, nous entreprenons un voyage complet à travers le paysage de la politique monétaire du Maroc, explorant son évolution historique, ses défis contemporains et ses perspectives futures. À travers une analyse rigoureuse, une réflexion critique et des perspectives prospectives, notre objectif est d'éclairer le chemin vers un avenir économique plus informé, résilient et prospère pour le Maroc et au-delà.

Dans les sections suivantes de ce rapport, nous explorerons en détail le contexte de notre projet de fin d'études, en commençant par une présentation de l'organisation et une analyse de la structure du système financier marocain. Nous examinerons ensuite le cadre de la politique monétaire, les instruments utilisés et les décisions récentes prises dans ce domaine, en évaluant leurs impacts. Ensuite, nous aborderons la méthodologie de travail que nous avons adoptée pour mener à bien ce projet, en décrivant les différentes étapes de l'ingénierie, du développement et du déploiement. Ainsi, nous allons aborder la revue des technologies, outils et techniques utilisés, suivie d'une présentation du cadre conceptuel et de la méthodologie de collecte et de traitement des données. Nous aborderons aussi l'entraînement et l'évaluation des modèles d'apprentissage automatique, avant de présenter l'interface utilisateur et le tableau de bord développé pour visualiser les résultats de notre analyse. Enfin, nous concluons en résumant les principaux résultats obtenus, les défis rencontrés, et en proposant des recommandations pour les futures améliorations de ce projet.

Chapitre I : Contexte du PFE

1. Introduction

Ce chapitre offre une vue d'ensemble du contexte dans lequel s'inscrit notre Projet de Fin d'Études (PFE). Nous commençons par une présentation de l'organisation dans laquelle le stage a été effectué, suivie d'une analyse de la structure du système financier marocain. Ensuite, nous examinons le cadre de la politique monétaire au Maroc, en mettant en lumière les principaux acteurs, leurs rôles et les objectifs poursuivis. Nous explorons également les instruments de la politique monétaire utilisés par la Banque centrale du Maroc, ainsi que les décisions récentes prises et leurs impacts sur l'économie nationale. Ensuite, nous décrivons le cadre général du stage, en mettant en évidence la thématique abordée et les objectifs à atteindre. Enfin, nous détaillons le travail demandé dans le cadre du PFE et présentons la planification du projet, soulignant les différentes étapes et les délais prévus pour chaque phase.

2. Organisation d'accueil

2.1. Présentation de l'organisation



Bank Al-Maghrib [1] , érigée en tant que Banque centrale du Maroc, se distingue nettement des institutions bancaires traditionnelles. Contrairement à une banque de détail où le grand public peut effectuer des opérations courantes telles que l'ouverture de comptes, les virements, ou l'obtention de crédits, une banque centrale, telle que Bank Al-Maghrib, revêt le statut de "banque des banques". Chaque banque commerciale du pays y détient un compte, soumis à des provisions obligatoires. Souvent désignée comme la "Banque nationale" ou le "Banquier de l'État", elle est parfois affublée du surnom de la "vieille Dame" dans certaines régions. Fondée en 1959 en remplacement de l'ancienne "Banque d'État du Maroc", son siège est stratégiquement situé à Rabat. L'institution étend son influence à travers le royaume avec deux succursales à Rabat et Casablanca, ainsi que 18 agences déployées de manière stratégique sur l'ensemble du territoire marocain.

2.2. Fonctions et missions

Parmi les fonctions de Bank Al-Maghrib, il y a :

- La fabrication de la monnaie fiduciaire et la préservation de sa sécurité
- L'élaboration et la mise en œuvre de la politique monétaire
- La gestion des réserves de change
- La supervision du système bancaire
- La sécurisation des systèmes et moyens de paiement

Ce n'est pas tout. Bank Al-Maghrib assure un grand nombre d'activités telles que la préservation de la valeur du dirham, l'élaboration et la publication des statistiques monétaires du pays, la production d'un certain nombre de rapports comme le rapport remis annuellement à Sa Majesté le Roi sans oublier la contribution au renforcement de la stabilité financière au niveau national.

2.3. Organigramme de l'organisation

L'organigramme de l'organisation constitue une représentation graphique de la structure hiérarchique et fonctionnelle de l'entité dans laquelle se déroule le Projet de Fin d'Études (PFE). Cette section offre un aperçu clair des différents départements, des relations hiérarchiques entre les membres du personnel, ainsi que des responsabilités spécifiques de chaque unité au sein de l'organisation. En examinant l'organigramme, les lecteurs peuvent comprendre la répartition des tâches, les flux de communication et les niveaux de supervision existants au sein de l'entité, ce qui est essentiel pour appréhender le contexte dans lequel le PFE est mené.

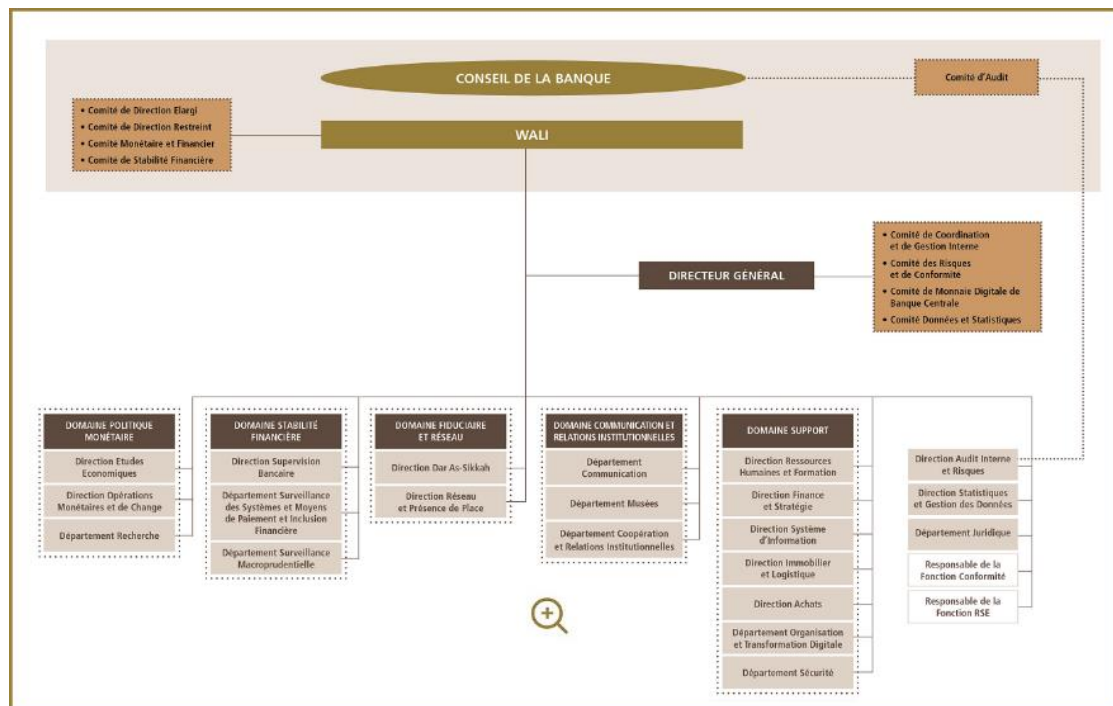


Figure 1: Organigramme de Bkam

3. Cadre de la politique monétaire

3.1. Structure du système financier marocain

La structure du système financier marocain est influencée par des réformes continues visant à renforcer la stabilité, la transparence et à favoriser l'innovation financière [2]. La modernisation du système financier s'inscrit dans le cadre des efforts du Maroc pour attirer les investissements et promouvoir la croissance économique. Voici une vue d'ensemble de la structure du système financier marocain [3] :

Tableau 1: structure du système financier marocain

Institutions financières	Banques	Le secteur bancaire marocain est dominé par plusieurs grandes banques, dont Attijariwafa Bank, Banque Populaire, et BMCE Bank. Il comprend également des banques étrangères opérant au Maroc.	
	Sociétés d'assurance	Le marché de l'assurance au Maroc est représenté par plusieurs compagnies d'assurance, offrant une gamme de produits d'assurance vie et non-vie	
	Microfinance	Des institutions de microfinance opèrent pour soutenir l'inclusion financière et offrir des services aux populations à faible revenu	
	Sociétés de financement	Les sociétés de financement sont des établissements financiers agréés par Bank Al-Magrib qui proposent des produits et services financiers autres que la collecte de dépôts. Elles peuvent octroyer des crédits aux particuliers et aux entreprises, proposer des opérations de leasing, du factoring ou du financement de projets d'investissement. Exemple : Maroc Lease	
	Associations de micro-crédit	Les associations de micro-crédit sont des institutions à vocation sociale qui accordent de petits prêts aux entrepreneurs et aux personnes à faibles revenus pour financer leurs activités et favoriser leur inclusion financière. Elles jouent un rôle important dans le développement économique et social du Maroc. Exemple : Aman Iman	
	Banques offshores	Le Maroc ne dispose pas de banques offshores à proprement parler. Les banques offshores sont des établissements bancaires situés dans des zones franches fiscales et qui proposent des services financiers à une clientèle internationale en dehors du cadre réglementaire national strict.	
	Banques participatives	Les banques participatives sont des banques islamiques qui opèrent selon les principes de la finance islamique. Elles ne pratiquent pas l'intérêt, mais se basent sur des modes de financement participatifs tels que la Mourabaha (vente à profit) ou la Musharaka (entreprise commune). Le Maroc a connu une croissance significative du secteur de la banque participative ces dernières années. Exemple : Bank Al Maghrib participatif	
Marchés financiers	Bourse de Casablanca	La Bourse de Casablanca est le principal marché boursier au Maroc. Elle facilite la négociation d'actions, d'obligations et d'autres instruments financiers	
	Marché monétaire et obligataire	Le Maroc dispose d'un marché monétaire et obligataire développé pour les transactions à court et à long terme	
Instruments financiers	Actions	Actions ordinaires	Confèrent à leur détenteur la propriété d'une fraction du capital de la société, le droit de vote aux assemblées générales et le droit aux dividendes.
		Actions de préférence	Offrent un droit de priorité sur les dividendes et/ou le remboursement du capital en cas de liquidation de la société, mais ne donnent pas toujours le droit de vote.
		Certificats d'investissement	Titres de créances négociables émis par les sociétés anonymes et qui donnent droit à un intérêt fixe.
		Parts de sociétés civiles immobilières (SCI)	Représentent une fraction de la propriété d'un immeuble détenu par une SCI.
	Obligations	Obligations ordinaires	Titres de créances négociables émis par les sociétés ou l'État qui donnent droit à un intérêt fixe et au remboursement du capital à une échéance déterminée.
		Obligations convertibles	Offrent la possibilité à leur détenteur de les convertir en actions de la société émettrice à une date et un prix déterminé.

		Obligations remboursables en actions	Offrent à leur détenteur le remboursement du capital en actions de la société émettrice à une échéance déterminée.
		Obligations à taux variable	Le taux d'intérêt de ces obligations est variable et peut être indexé sur un indice de référence comme le taux Libor.
		Obligations à coupon zéro	Ne comportent pas de coupon d'intérêt et sont émises à un prix décoté de leur valeur nominale.
	Titres de créances	Billets de trésorerie	Titres de créances à court terme émis par le Trésor public. Exemple : Billets de trésorerie émis par la Direction Générale du Trésor et de la Comptabilité Publique
		Certificats de dépôt	Titres de créances à court terme émis par les banques et qui donnent droit à un intérêt rémunérateur. Exemple : Certificats de dépôt à la Banque Centrale Populaire
		Bons du Trésor	Titres de créances à moyen et long terme émis par le Trésor public.
		Sukuk	Certificats d'investissement islamiques qui représentent une part d'un actif tangible ou d'un projet et qui génèrent des revenus locatifs ou des profits commerciaux.
	Dérivés	Futures	Contrats standardisés qui obligent les parties à acheter ou à vendre un actif sous-jacent à une date et un prix déterminé. Exemple : Contrats à terme sur le blé
		Options	Contrats qui donnent à leur détenteur le droit, mais non l'obligation, d'acheter ou de vendre un actif sous-jacent à une date et un prix déterminé. Exemple : Options d'achat sur l'action
Régulateurs financiers	Banque Al-Maghrib	La Banque centrale du Maroc est responsable de la régulation monétaire et bancaire.	
	Autorité Marocaine du Marché des Capitaux (AMMC)	Régulateur des marchés financiers	
Infrastructure financière	Maroclear	Le dépositaire central au Maroc, facilitant la conservation des titres	
	Système de paiement	Les infrastructures permettent des transactions financières efficaces	
Services financiers	Services de microfinance	Objectif : Offrir des services financiers de base tels que des prêts, des dépôts et des transferts d'argent aux populations à faibles revenus et exclues du système bancaire traditionnel. Institutions : Associations de micro-crédit, institutions de microfinance, banques participatives. Exemple : Accès aux services financiers pour les micro-entrepreneurs et les artisans.	
	Services de planification de la retraite	Objectif : Aider les individus à épargner et à investir pour leur retraite afin de garantir un revenu stable et une qualité de vie décente après la cessation d'activité. Produits : Plans d'épargne retraite, contrats d'assurance-vie, fonds de pension. Exemple : Adhésion à un plan d'épargne retraite individuel ou collectif.	

	Finance islamique	Principes : Basée sur les principes de la Charia islamique, qui interdisent l'usure et encouragent le partage des profits et des pertes. Produits : Sukuk, Mourabaha, Musharaka, Takaful (assurance islamique). Exemple : Financement d'un achat immobilier via un prêt immobilier islamique.
	Services de transfert d'argent	Objectif : Faciliter l'envoi et la réception d'argent entre des personnes situées dans différentes régions ou pays. Opérateurs: Western Union, MoneyGram, Wafacash, Barid Cash. Exemple : Envoi d'argent à un membre de la famille vivant à l'étranger.
	Assurances	Les compagnies d'assurance fournissent des services de protection
Politiques monétaires et fiscales	Banque Al-Maghrib	Responsable de la politique monétaire
	Ministère de l'Économie et des Finances	Impliqué dans les politiques fiscales

3.2. La politique monétaire

Le système financier marocain a subi des réformes majeures depuis les années 90, se concentrant sur le décloisonnement des marchés de capitaux, la libéralisation des opérations financières [2] , et la réforme réglementaire des banques et du marché financier. Les réformes, notamment la loi bancaire de 1993, ont introduit le désencadrement du crédit, la libéralisation des taux d'intérêt, et ont renforcé le cadre prudentiel. Malgré les progrès, la recherche de taille critique pour concurrencer les grandes banques d'autres pays pose des défis, soulevant des questions sur les modalités de concentration. En 2011, le système financier marocain est considéré comme le plus structuré de la rive sud de la Méditerranée, mais la consolidation reste un enjeu.

La politique économique, dirigée par des recommandations d'économistes, est une politique conjoncturelle qui agit sur la demande globale (consommation, investissement) pour réduire les écarts par rapport à l'équilibre. Elle vise à stimuler les revenus, maintenir des taux d'inflation et de chômage bas, et atteindre un équilibre dans les soldes publics et extérieurs. La politique monétaire [4] , centrée sur la tension entre inflation et chômage, consiste en des décisions de l'autorité ou de la banque centrale pour contrôler la quantité de monnaie en circulation. Elle affecte les taux d'intérêt et utilise divers instruments tels que les opérations sur le marché, les réserves obligatoires, et les taux directeurs [5] .

Parmi ces instruments, les taux directeurs, décidés par la banque centrale, sont cruciaux. Des taux élevés minimisent la demande, réduisent l'inflation et le chômage, tandis que des taux bas stimulent la demande, augmentent le revenu et ralentissent le chômage, mais peuvent provoquer une inflation à long terme.

La banque centrale contrôle la masse monétaire, emprunte aux banques si nécessaire, et fixe les conditions d'emprunt via les taux directeurs. Le taux interbancaire, influencé par le taux directeur, varie quotidiennement et impacte les coûts d'emprunt à court terme (crédit) et à long terme (investissement).

Les taux d'intérêt élevés découragent les investissements, rendent l'emprunt plus cher pour les consommateurs et peuvent attirer des capitaux étrangers, provoquant une appréciation de la devise. À l'inverse, des taux bas encouragent les investissements et la consommation, mais peuvent déprécier la devise. [6]

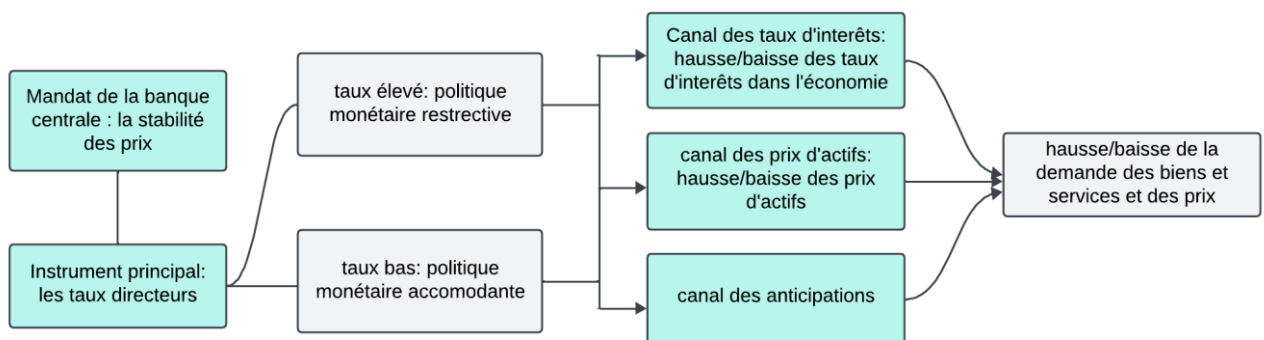


Figure 2: l'effet de la politique monétaire

3.3. Les instruments de la politique monétaire

Les instruments de la politique monétaire sont les outils utilisés par la Banque centrale pour influencer la masse monétaire, les taux d'intérêt, le contrôle de l'inflation, la stabilité des prix, le plein emploi et la croissance économique. Voici une description détaillée de certains de ces instruments :

- **Taux d'intérêt directeurs** : Les taux d'intérêt directeurs sont des taux d'intérêt fixés par une banque centrale, comme la BAM ou la Réserve fédérale américaine (Fed), qui influencent le niveau des taux d'intérêt dans l'ensemble de l'économie. Ils sont utilisés comme outils de politique monétaire pour réguler la croissance économique et l'inflation. Voici les principaux taux d'intérêt directeurs :
 - **Taux de refinancement (ou taux repo)** : C'est le taux auquel les banques commerciales peuvent emprunter des fonds à court terme auprès de la banque centrale. Lorsque ce taux est bas, les emprunts sont moins coûteux, ce qui encourage les banques à prêter davantage aux entreprises et aux consommateurs, stimulant ainsi l'économie. À l'inverse, un taux de refinancement élevé rend les emprunts plus coûteux, ce qui peut freiner l'inflation en réduisant la demande de crédit.
 - **Taux d'escompte** : C'est le taux auquel les banques commerciales peuvent emprunter directement auprès de la Banque centrale en cas de besoin. En ajustant ce taux, la Banque centrale peut influencer les coûts d'emprunt dans l'ensemble du système financier.

- **Taux des dépôts** : C'est le taux d'intérêt que la banque centrale paie sur les dépôts que les banques commerciales y placent. Un taux des dépôts élevé peut inciter les banques à conserver des réserves excédentaires à la banque centrale plutôt que de les prêter, réduisant ainsi la liquidité dans l'économie.
- **Taux des opérations de marché ouvert (OMO)** : La Banque centrale peut acheter ou vendre des titres sur le marché ouvert pour influencer la quantité de monnaie en circulation. Lorsqu'elle achète des titres, elle injecte de l'argent dans l'économie, et lorsqu'elle vend des titres, elle retire de l'argent.
- **Réserves obligatoires** : Les banques commerciales sont tenues de déposer une certaine quantité de leurs réserves à la Banque centrale. En ajustant le niveau de ces réserves obligatoires, la Banque centrale peut contrôler la quantité de monnaie que les banques peuvent prêter.
- **Politique des taux planchers ou plafonds** : La Banque centrale peut définir des taux planchers ou plafonds pour les taux d'intérêt sur le marché interbancaire. Cela influence les taux d'intérêt à travers l'économie. Par exemple, si elle fixe un taux plancher, les taux d'intérêt ne devraient pas descendre en dessous de ce seuil.
- **Communication et guidance** : La Banque centrale utilise souvent la communication pour orienter les attentes du marché. Par exemple, elle peut émettre des déclarations indiquant ses intentions concernant les taux d'intérêt futurs, ce qui influence les décisions d'emprunt et d'investissement.
- **Opérations de pension** : Ces opérations impliquent l'achat ou la vente de titres financiers avec un engagement de les racheter ou de les revendre à une date ultérieure. Cela influence les taux d'intérêt à court terme.
- **Contrôle macroprudentiel** : La Banque centrale peut mettre en œuvre des mesures macroprudentielles, telles que des exigences de fonds propres plus élevées pour les prêts immobiliers, pour atténuer les risques systémiques liés à certaines activités financières.
- **Politique de taux de change** : Dans les pays ayant une monnaie convertible, la Banque centrale peut intervenir sur le marché des changes pour influencer la valeur de la monnaie nationale.
- **Instruments non conventionnels** : En période de crise, la Banque centrale peut utiliser des instruments non conventionnels tels que l'assouplissement quantitatif (achats massifs d'actifs) pour stimuler l'économie lorsque les taux d'intérêt traditionnels sont proches de zéro.

3.4. Décisions récentes de politique monétaire et leurs impacts

Réunie le 19 décembre 2023, Bank Al-Maghrib a pris plusieurs décisions de politique monétaire cruciales pour l'économie marocaine [6]. La banque centrale a choisi de maintenir son taux directeur inchangé à 3%, indiquant ainsi une volonté de soutenir l'activité économique tout en surveillant l'évolution de l'inflation. Cette décision intervient dans un contexte mondial marqué par une atténuation des pressions inflationnistes et une décélération de la croissance économique. Malgré les défis liés au séisme d'Al Haouz, les perspectives à moyen et long terme restent optimistes grâce à des projets d'investissement d'envergure. En adoptant une approche prudente, Bank Al-Maghrib vise à maintenir la stabilité des prix, avec des projections de décélération de l'inflation à moyen terme. Les taux débiteurs ont également augmenté, reflétant le resserrement de la politique monétaire. Ces

décisions visent à équilibrer la croissance économique, la stabilité des prix et la gestion des pressions inflationnistes, contribuant ainsi à façonner un environnement économique durable pour le Maroc.

4. Cadre général du stage

Le stage que j'entreprends s'inscrit dans le cadre d'un projet d'analyse des retombées de presse sur la politique monétaire du Maroc. Ce projet vise à comprendre les opinions et les sentiments du public à l'égard des décisions de la banque centrale marocaine, Bank Al-Maghrib, ainsi que leur évolution en fonction des conditions économiques et des communications de cette institution. L'objectif principal est d'utiliser l'analyse de sentiment pour améliorer la communication et la prise de décision de Bank Al-Maghrib, en s'appuyant sur un corpus de données en langue française spécifiquement dédié à la politique monétaire du pays.

Dans ce contexte, le stage comprend plusieurs volets. Tout d'abord, il implique la collecte de données pertinentes en français à partir des articles de presse. Ensuite, il nécessite la construction d'un corpus de données représentatif et équilibré, reflétant les différentes perspectives et opinions du public marocain sur la politique monétaire.

Par la suite, le stage comprend également une phase d'entraînement des modèles d'apprentissage automatique pour l'analyse de sentiment. Cette étape consiste à explorer différentes architectures de modèles, techniques de prétraitement du texte et paramètres d'entraînement afin d'optimiser les performances des modèles.

Un autre volet important du stage est la création d'un tableau de bord interactif. Ce tableau de bord sera conçu pour visualiser les résultats de l'analyse de sentiment de manière intuitive et accessible. Il comprendra des graphiques, des nuages de mots, des statistiques détaillées et une section dédiée à la reconnaissance des entités nommées (NER), permettant ainsi aux utilisateurs de naviguer facilement à travers les données et d'interpréter les résultats.

Enfin, le stage comprendra une évaluation approfondie des modèles entraînés, ainsi qu'une analyse des résultats obtenus. Cette analyse permettra de comprendre les forces et les faiblesses des différents modèles, et servira de base à des recommandations pour l'amélioration continue du projet.

Dans l'ensemble, le cadre général du stage comprend donc la collecte et la construction d'un corpus de données, l'entraînement des modèles d'apprentissage automatique, et l'évaluation des performances de ces modèles. Ces activités seront menées dans le but ultime d'améliorer la communication et la prise de décision de Bank Al-Maghrib en matière de politique monétaire, contribuant ainsi à une meilleure compréhension des perceptions du public et des acteurs économiques au Maroc.

5. Thématique

5.1. Présentation du sujet

Le thème de ce stage tourne autour d'établir une analyse des sentiments dans les textes relatifs aux décisions de politique monétaire marocaine. Ce thème encapsule une exploration multifacette du discours économique, du traitement automatique du langage naturel (TALN) et de l'analyse des sentiments dans le contexte spécifique de l'élaboration des politiques économiques marocaines.

Au cœur du projet, il s'agit de répondre au besoin d'obtenir des informations complètes et fondées sur les données concernant les sentiments, les perceptions et les implications présents dans les sources textuelles pertinentes aux décisions de politique monétaire marocaine. En construisant un corpus spécialisé comprenant une gamme diversifiée de données textuelles, y compris des articles de presse, des déclarations politiques, des rapports économiques et d'autres sources pertinentes, le stage vise à créer un riche répertoire de données linguistiques pour l'analyse, servant de moyen essentiel pour atteindre l'objectif principal.

À travers le prisme de l'analyse des sentiments, le projet vise à découvrir les sentiments sous-jacents et les attitudes envers divers indicateurs économiques, les décisions politiques et leurs impacts potentiels sur l'économie marocaine. En appliquant des techniques avancées en traitement automatique du langage naturel, en apprentissage automatique et en analyse des sentiments, le stage s'efforce d'extraire des informations précieuses des sources textuelles, discernant le ton, la polarité et les nuances contextuelles présents dans le langage utilisé pour discuter des questions de politique monétaire.

De plus, le thème du stage souligne l'importance plus large de comprendre et d'analyser le discours économique dans le contexte de l'élaboration des politiques économiques marocaines. Alors que le Maroc navigue à travers des paysages économiques complexes caractérisés par la volatilité, l'incertitude et l'évolution des dynamiques mondiales, le projet vise à fournir une approche fondée sur les données pour décrypter les tendances économiques, le sentiment du marché et les perceptions des parties prenantes.

5.2. Travail demandé

Ce stage se déroulera au sein de la Direction des Statistiques et de la Gestion des Données de Bank Al-Maghrib à partir du février 2024. Durant cette période, les tâches suivantes seront réalisées :

- Collecte des données issues d'articles de presse en français portant sur les récentes décisions de politique monétaire dans la période de 2020 à 2024
- Nettoyage et traitement des données collectées.
- Étiquetage des données et création d'un premier corpus en français.
- Mise en place d'un modèle de traitement automatique du langage naturel (NLP) pour obtenir

les premiers résultats.

- Entraînement du corpus et du modèle NLP.
- Création d'un tableau de bord pour l'analyse des sentiments

5.3. Délimitations

Le projet est délimité selon les critères suivants :

- **Période d'étude** : Le projet se concentre sur la période allant de 2020 à 2024, permettant ainsi d'analyser les répercussions des décisions récentes de politique monétaire sur la presse au Maroc.
- **Sources de données** : Les données utilisées seront exclusivement des articles de presse en ligne, offrant ainsi une base solide et actuelle pour l'analyse des réactions médiatiques aux décisions de politique monétaire.
- **Méthodes d'analyse** : L'analyse sera effectuée en utilisant des techniques d'analyse de sentiments lexicales et basées sur l'apprentissage automatique, permettant une évaluation approfondie des opinions exprimées dans les articles de presse concernant la politique monétaire.
- **Outils et technologies** : Le projet fera usage d'outils et de technologies modernes tels que Python, spaCy et NLTK, reconnus pour leur efficacité dans le traitement du langage naturel et l'analyse de données. Cela garantira une approche robuste et précise pour l'analyse des sentiments dans la presse.

5.4. Objectifs à atteindre

L'objectif principal du projet est de développer un corpus et un algorithme d'analyse de sentiments pour l'analyse des retombées de la presse sur les décisions de politique monétaire au Maroc. Cela implique la création d'une base de données comprenant des articles de presse pertinents et leur annotation en termes de sentiments exprimés à l'égard de la politique monétaire.

Les objectifs secondaires du projet sont les suivants :

- Identifier les thématiques et les opinions dominantes dans la presse : En analysant les articles de presse, il s'agit de repérer les sujets récurrents et les tendances générales dans les opinions exprimées concernant la politique monétaire. Cela permettra de mieux comprendre les préoccupations et les perceptions des médias sur ce sujet.
- Mesurer l'impact des décisions de politique monétaire sur l'opinion publique : En évaluant les sentiments exprimés dans la presse suite aux décisions de politique monétaire, le projet vise à quantifier l'impact de ces décisions sur l'opinion publique. Cela permettra d'évaluer dans quelle mesure les actions de la Banque centrale sont perçues positivement ou négativement

par les médias et, par extension, par le grand public.

- Améliorer la communication de la Banque sur ses décisions : En comprenant mieux les réactions de la presse aux décisions de politique monétaire, le projet pourrait fournir des indications précieuses sur la manière dont la Banque centrale communique ses actions et ses objectifs. Cela pourrait conduire à des ajustements dans la stratégie de communication pour mieux informer et influencer l'opinion publique.

5.5. Planification du projet

Au cours des réunions avec notre superviseur de stage, nous avons entamé la planification et la définition des objectifs de notre projet. Nous avons identifié clairement ce que nous souhaitons réaliser avec le corpus et les résultats attendus de l'analyse de sentiment. En parallèle, nous avons déterminé les exigences spécifiques en termes de données, de modèles d'apprentissage automatique et de technologies à utiliser pour atteindre ces objectifs.

Par la suite, nous allons procéder à la collecte et au prétraitement des données en français, en puisant dans diverses sources telles que les articles de presse et les discours officiels. Ces données seront prétraitées pour éliminer le bruit, normaliser le texte et effectuer d'autres transformations afin d'assurer la qualité des données. Parallèlement, nous allons entreprendre la construction d'un corpus de données en français spécifiquement dédié à la politique monétaire du Maroc, en veillant à ce qu'il soit représentatif et équilibré pour capturer les différentes perspectives et opinions du public. Ensuite, nous allons procéder à l'entraînement des modèles d'apprentissage automatique pour l'analyse de sentiment en utilisant les données du corpus construit. Nous allons explorer différentes architectures de modèles, techniques de prétraitement du texte et paramètres d'entraînement pour optimiser les performances des modèles. Ainsi, nous allons évaluer les performances des modèles entraînés en utilisant des mesures telles que la précision, le rappel et la F-mesure, tout en effectuant une analyse approfondie des résultats pour comprendre les forces et les faiblesses des différents modèles.

Enfin, nous prévoyons de développer un tableau de bord pour visualiser les résultats de l'analyse de sentiment. Ce tableau de bord permettra de présenter de manière claire et concise les informations clés extraites des textes analysés, telles que les tendances des sentiments, les mots les plus couramment utilisés, et les entités nommées détectées. Cela offrira à Bank Al-Maghrib une interface interactive pour explorer les perceptions du public et les aidera à prendre des décisions éclairées basées sur les données collectées.

Le diagramme de Gantt représente de manière visuelle les différentes étapes et les délais prévus pour la réalisation de mon projet d'analyse des retombées médiatiques sur la politique monétaire du Maroc. Ce planning détaillé offre une vue d'ensemble du déroulement du projet, depuis la collecte des données jusqu'à l'évaluation des modèles d'analyse de sentiment. En organisant les tâches par séquence et en assignant des durées spécifiques à chacune, le diagramme de Gantt permet une gestion efficace du temps et des ressources, tout en assurant une progression méthodique vers l'atteinte des

objectifs fixés.

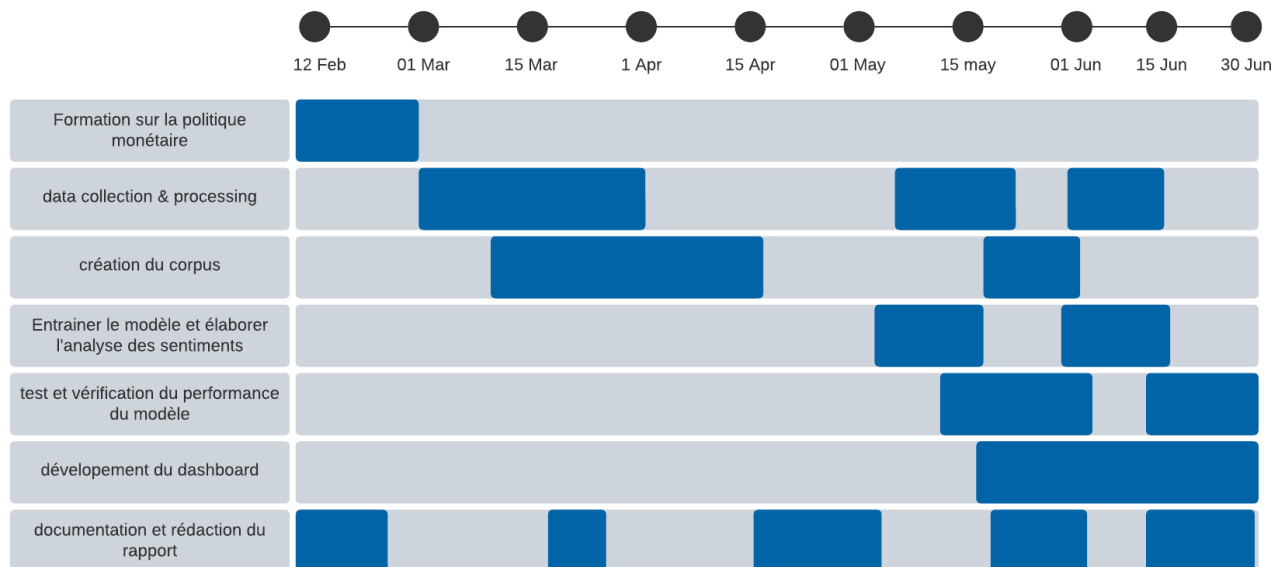


Figure 3: Diagramme de Gantt

6. Conclusion

Ce chapitre a permis de dresser un panorama complet du contexte dans lequel s'inscrit le Projet de Fin d'Études (PFE). À travers la présentation de l'organisation d'accueil et de la structure du système financier marocain, nous avons pu poser les bases nécessaires à la compréhension des enjeux économiques et monétaires du pays. En examinant le cadre de la politique monétaire, nous avons identifié les acteurs clés et les instruments utilisés pour réguler l'économie. De plus, l'analyse des décisions récentes de politique monétaire a permis de saisir l'impact de ces mesures sur l'économie nationale. En détaillant le cadre général du stage, nous avons précisé la thématique abordée et les objectifs visés. Enfin, la planification du projet a été établie, définissant les étapes à suivre pour mener à bien le PFE. Ce chapitre jette ainsi les bases nécessaires à la compréhension approfondie de la problématique et de l'analyse qui seront abordées dans les chapitres suivants.

Chapitre II : Problématique et analyse

1. Introduction

Ce chapitre examinera la solution proposée après avoir analysé les données et les méthodes existantes pour la construction du corpus et l'analyse des sentiments. Il présentera également les objectifs de l'étude, ainsi qu'une analyse des besoins fonctionnels et non fonctionnels associés à la collecte et à l'analyse des données textuelles.

2. Problématique et positionnement

2.1. Contexte général

Le contexte financier marocain a été profondément transformé depuis les années 90, avec des réformes axées sur la libéralisation des marchés de capitaux, la réglementation financière et bancaire, et la modernisation des pratiques économiques. Ces changements, notamment initiés par la loi bancaire de 1993, ont permis une plus grande flexibilité dans le système financier, favorisant la concurrence et la croissance économique. Malgré ces progrès, les défis persistent, notamment en ce qui concerne la concentration du secteur financier et la nécessité de maintenir une croissance stable et durable.

La politique économique marocaine, guidée par des recommandations d'experts, vise à stimuler la croissance, à maintenir des niveaux d'inflation et de chômage bas, et à assurer l'équilibre des comptes publics et extérieurs. La politique monétaire, en particulier, est un levier crucial pour atteindre ces objectifs, en ajustant les taux d'intérêt et en contrôlant la masse monétaire pour influencer l'activité économique. Les décisions de la banque centrale, telles que les variations des taux directeurs, ont un impact significatif sur les conditions de prêt, l'investissement et la consommation dans l'économie.

Bank Al-Maghrib, en tant qu'institution centrale de la politique monétaire, joue un rôle essentiel dans la stabilité financière et la croissance économique du Maroc. En fixant les taux d'intérêt directeurs et en utilisant d'autres instruments de politique monétaire, elle cherche à maintenir un équilibre entre la croissance et la stabilité des prix. Les décisions récentes de Bank Al-Maghrib, prises en décembre 2023, reflètent cette volonté de soutenir l'activité économique tout en surveillant l'inflation, dans un contexte de pressions économiques mondiales et de défis internes.

Les implications de la politique monétaire vont au-delà des marchés financiers, affectant les entreprises, les ménages et les investisseurs. Les variations des taux d'intérêt peuvent influencer les décisions d'emprunt et d'investissement, tandis que la communication de la banque centrale joue un rôle crucial dans la gestion des attentes et la stabilité économique. Une communication claire et cohérente contribue à réduire l'incertitude et à favoriser la confiance des acteurs économiques, créant ainsi un environnement propice à la croissance et à la prospérité économique.

2.2. Problématique

La problématique de cette étude se décline en plusieurs questions clés. Tout d'abord, nous cherchons à comprendre les opinions et les sentiments du public à l'égard de la politique monétaire de Bank Al-Maghrib. Nous nous intéressons également à l'évolution de ces opinions et sentiments en fonction des décisions de la banque centrale et des conditions économiques. De plus, nous nous interrogeons sur les facteurs qui influencent ces opinions et sentiments, afin de mieux cerner les dynamiques socio-économiques en jeu. Enfin, nous nous demandons comment Bank Al-Maghrib pourrait tirer parti de l'analyse de sentiment pour améliorer sa communication et sa prise de décision. Face à ces questions, plusieurs défis se présentent. Tout d'abord, l'absence d'un corpus de données en langue française spécifiquement dédié à la politique monétaire du Maroc constitue un obstacle majeur pour mener des analyses approfondies et précises. Ensuite, la complexité d'analyser les retombées de la presse sur les décisions de politique monétaire pose un défi supplémentaire, nécessitant des efforts considérables en raison de la quantité de données et de leur nature subjective. Enfin, il est impératif de développer un outil automatisé pour l'analyse des sentiments afin de faciliter le processus d'évaluation des réactions du public et des acteurs économiques, offrant ainsi des informations précieuses pour les décideurs politiques et les analystes financiers.

2.3. Analyse de l'existant

Dans le domaine de l'analyse des retombées de la politique monétaire au Maroc, plusieurs méthodes traditionnelles ont été utilisées pour évaluer les opinions et les sentiments du public ainsi que les répercussions des décisions de Bank Al-Maghrib. Les données macroéconomiques, telles que le PIB, l'inflation et le chômage, ont longtemps été des indicateurs essentiels pour évaluer l'état de l'économie et orienter les décisions de politique monétaire. De plus, les enquêtes menées auprès des entreprises, des consommateurs et des investisseurs ont permis de recueillir des opinions subjectives sur l'économie et la politique monétaire. Enfin, l'utilisation de modèles économétriques a permis de simuler l'impact probable des décisions de politique monétaire sur l'économie.

Ces méthodes traditionnelles présentent des avantages, notamment leur solidité théorique et leur utilisation de données historiques fiables. Cependant, elles ont également des limites, telles que le délai dans la collecte des données, la nécessité de suppositions simplificatrices dans les modèles économétriques, et le fait qu'elles ne capturent pas toujours les nuances des réponses émotionnelles et subjectives des acteurs économiques.

Pour enrichir l'analyse de sentiments dans le contexte de la politique monétaire marocaine, plusieurs corpus existants peuvent être explorés. Le premier, "Corpus French Twitter Sentiment", est spécifiquement construit à partir de tweets en français et annoté avec des polarités de sentiment. Étant basé sur des données provenant de Twitter, il capture souvent un langage plus informel et des expressions courantes utilisées sur les réseaux sociaux, offrant ainsi une perspective unique sur les

opinions et les sentiments du public.

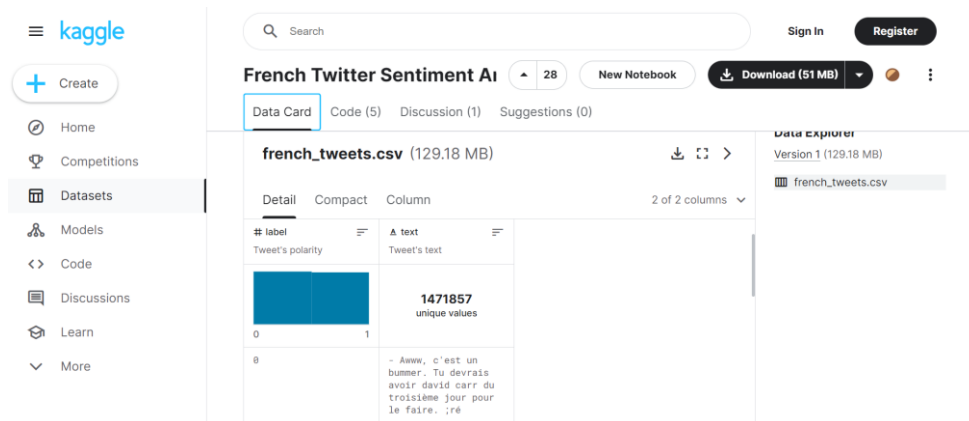


Figure 4: Kaggle- French corpus for sentiment analysis

En outre, le corpus Tatoeba, bien qu'il ne soit pas spécifiquement conçu pour l'analyse de sentiments, peut être utilisé pour enrichir la variété des textes utilisés dans l'analyse. Il s'agit d'une collection de phrases et de traductions dans de nombreuses langues, y compris le français. Bien que principalement utilisé pour la traduction et l'apprentissage des langues, ce corpus peut contenir des phrases pertinentes pour la politique monétaire en fonction du contexte.

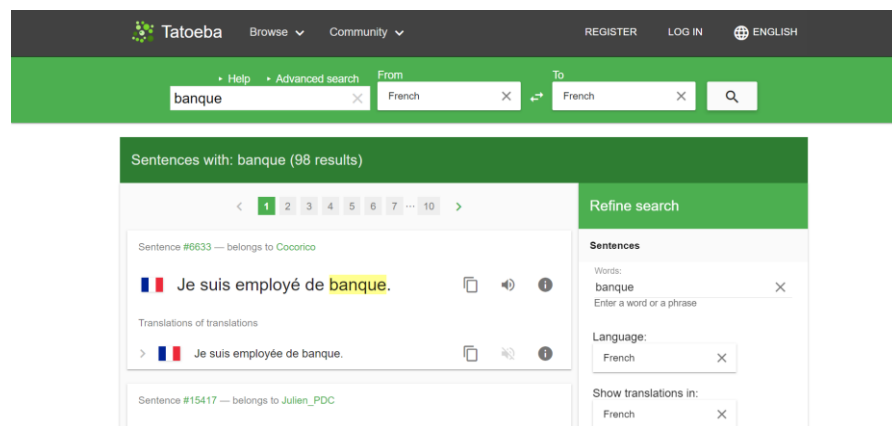


Figure 5: Tatoeba dictionary

Sketch Engine est un logiciel développé par Lexical Computing CZ s.r.o. depuis 2003, conçu pour gérer les corpus et analyser le texte. Son objectif principal est de faciliter l'étude du comportement linguistique, en répondant aux besoins des lexicographes, des chercheurs en linguistique de corpus, des traducteurs et des apprenants en langues. Cet outil permet aux utilisateurs de rechercher dans de vastes collections de texte en utilisant des requêtes sophistiquées et motivées linguistiquement. Une de ses fonctionnalités remarquables est les esquisses de mots, qui sont des résumés automatiques d'une page dérivés des corpus, montrant le comportement grammatical et collocationnel d'un mot. Sketch Engine prend en charge et offre des corpus dans plus de 90 langues, en faisant un outil polyvalent pour l'analyse linguistique et la recherche.

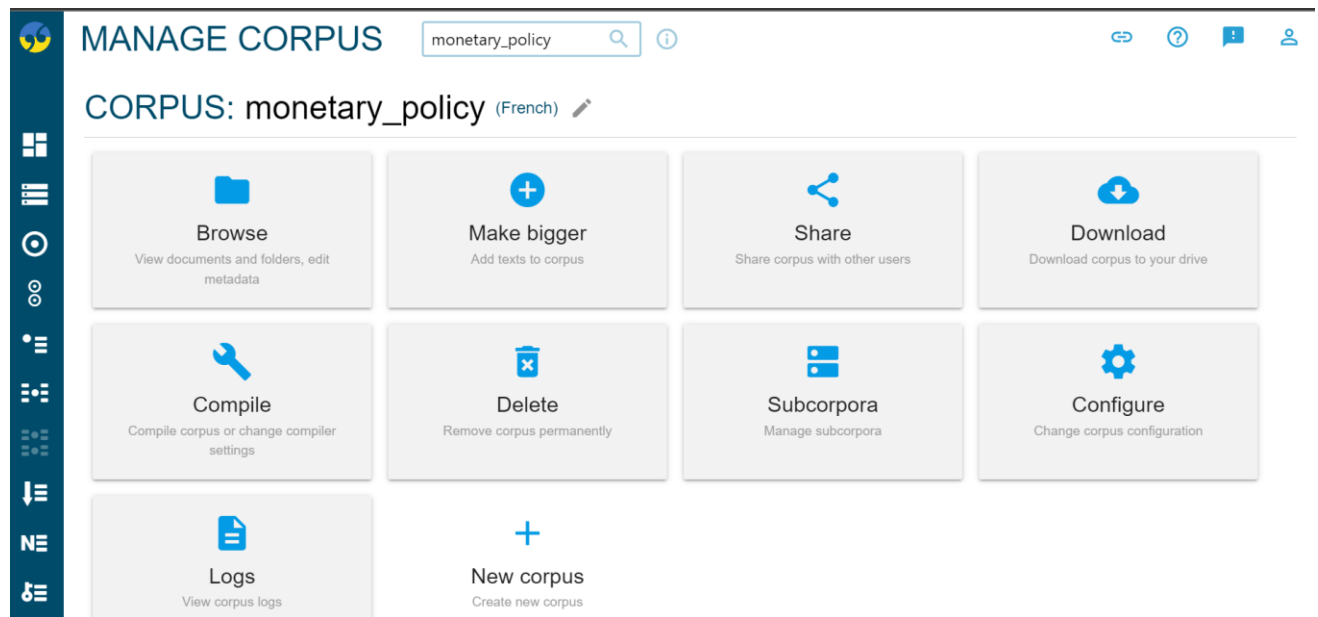


Figure 6: sketch engine - corpus maker

Ces corpus offrent une base de données diversifiée pour l'analyse de sentiments, fournissant ainsi des informations précieuses pour comprendre les tendances et les réactions du public à la politique monétaire au Maroc. Cependant, il est important de noter que ces corpus présentent également des limites, telles que le manque de données spécifiques à la politique monétaire marocaine et la nécessité d'annotations de sentiment précises pour des analyses plus fines.

2.4. Critique de l'existant

Malgré les avantages des méthodes traditionnelles, elles présentent des limites importantes dans le contexte de l'analyse des retombées de la politique monétaire au Maroc. Par exemple, les enquêtes peuvent être sujettes à des biais et ne captent pas toujours les réactions en temps réel du public aux événements économiques. De plus, les modèles économétriques reposent souvent sur des hypothèses simplificatrices et peuvent manquer de précision dans la prévision des réactions du marché à des événements spécifiques.

Par conséquent, il existe un besoin croissant de méthodes plus réactives et diversifiées pour analyser les retombées de la politique monétaire. L'analyse de sentiments, alimentée par l'utilisation de données massives provenant des médias sociaux, des forums en ligne et des articles de presse, offre une perspective en temps réel et une compréhension plus nuancée des réactions du public. En combinant des approches traditionnelles avec des techniques d'analyse de sentiments plus innovantes, il est possible d'obtenir une image plus complète et précise des retombées de la politique monétaire au Maroc, ce qui pourrait informer de manière plus efficace les décisions de Bank Al-Maghrib et les politiques économiques du pays.

Tableau 2: étude comparative des méthodes existantes

	French Twitter Sentiment	Tatoeba	Sketch engine
Avantages	Corpus French Twitter Sentiment : Étant basé sur des tweets en français, ce corpus reflète les opinions et les sentiments exprimés sur les réseaux sociaux, ce qui peut être utile pour capturer les tendances et les réactions instantanées du public.	Bien que principalement utilisé pour la traduction et l'apprentissage des langues, ce corpus contient des phrases en français avec leurs équivalents dans d'autres langues, ce qui peut être exploité pour enrichir la variété des textes utilisés dans l'analyse de sentiments.	Interface conviviale et intuitive, propose une large gamme d'outils d'analyse linguistique, permet de créer des corpora personnalisés dans plus de 90 langues
Inconvénients	Les tweets peuvent être sujets à des bruits et des ambiguïtés linguistiques, ce qui rend parfois difficile l'interprétation précise des sentiments exprimés, en particulier dans un contexte formel comme celui de la politique monétaire.	Bien que comportant des phrases en français, ce corpus peut manquer de contexte et de cohérence dans les textes, ce qui peut rendre son utilisation moins efficace pour l'analyse de sentiments précise et ciblée.	Les Corpora peuvent être limités en taille ou en couverture L'annotation automatique de certains corpora peut parfois être sujette à des erreurs Le coût d'accès à certaines fonctionnalités premium peut être prohibitif pour certains utilisateurs ou organisations.
Taille du corpus	Bien que le corpus de tweets soit généralement volumineux, sa taille peut varier et ne pas toujours être suffisante pour représenter la diversité des opinions sur des sujets spécifiques tels que la politique monétaire.	Ce corpus contient une grande quantité de phrases en français, mais sa taille peut être insuffisante pour des analyses approfondies nécessitant un grand nombre de textes contextuellement riches.	Peut varier considérablement en fonction de la langue et du domaine d'étude. Pour les langues et les domaines plus couramment étudiés, des corpora de grande taille sont souvent disponibles,
Thèmes abordés	Les tweets peuvent aborder divers sujets, mais leur pertinence directe pour la politique monétaire peut être limitée en raison de la nature informelle et souvent brève des messages sur Twitter.	Ce corpus n'est pas spécifiquement orienté vers des thèmes particuliers, mais il peut contenir des phrases pertinentes pour la politique monétaire en fonction du contexte.	Couvre une vaste gamme de thèmes linguistiques, allant de la syntaxe à la sémantique en passant par la lexicographie et la traduction. Abordent également divers domaines de spécialisation, tels que la médecine, le droit, l'informatique, etc., ce qui permet aux chercheurs d'explorer des questions spécifiques liées à leur domaine d'intérêt.
Qualité de l'annotation	L'annotation des tweets peut être plus subjective en raison de la nature brève et informelle des messages, ce qui peut rendre la qualité de l'annotation moins fiable pour des analyses précises.	La qualité de l'annotation dépendra de la précision des traductions et de la pertinence contextuelle des phrases, ce qui peut varier et rendre l'annotation moins fiable pour des analyses de sentiments spécifiques.	Peut varier en fonction de la source des données et du processus d'annotation utilisé. Dans certains cas, l'annotation peut être effectuée automatiquement à l'aide d'algorithmes de traitement du langage naturel, ce qui peut entraîner des erreurs ou des incohérences dans les résultats.

3. Solutions proposées

Pour répondre aux besoins de comprendre les perceptions et les réactions du public envers la politique monétaire de Bank Al-Maghrib, une solution proposée consiste à développer un corpus de données en langue française spécifiquement dédié à cet objectif. Ce corpus permettrait d'analyser les articles de presse et les commentaires en français pour saisir les nuances et les subtilités de l'opinion publique. En identifiant les points de satisfaction, d'insatisfaction et les préoccupations du public, ainsi qu'en cernant les attentes et les besoins de la population, une communication plus ciblée et efficace de la part de la banque centrale pourrait être élaborée.

Pour renforcer la recherche et l'analyse économiques, on propose d'étudier l'influence des médias et des discours publics sur la formation de l'opinion publique et la prise de décision économique. En identifiant les biais et les facteurs influençant la perception du public envers la politique monétaire, des modèles d'analyse plus précis pourraient être développés pour prédire les réactions du public et les impacts économiques des décisions de la banque centrale.

Ainsi, on peut conclure l'importance d'un corpus dans les points suivants :

Tableau 3: l'importance d'un corpus dans le contexte de ce projet

Comprendre les perceptions et les réactions du public	Analyser les articles de presse et les commentaires en français permet de saisir les nuances et les subtilités de l'opinion publique envers la politique monétaire de Bank Al-Maghrib.
	Identifier les points de satisfaction, d'insatisfaction et les préoccupations du public concernant les décisions et leurs impacts économiques.
	Mieux cerner les attentes et les besoins de la population pour une communication plus ciblée et efficace de la part de la banque centrale.
Améliorer la communication et la prise de décision	Évaluer l'impact des communiqués de presse et des discours de Bank Al-Maghrib sur la perception du public et ajuster la stratégie de communication en conséquence.
	Identifier les points de confusion ou de mécompréhension pour les clarifier dans les communications futures.
	Anticiper les réactions potentielles aux nouvelles décisions de politique monétaire et adapter les mesures pour minimiser les impacts négatifs.
Renforcer la recherche et l'analyse économiques	Étudier l'influence des médias et des discours publics sur la formation de l'opinion publique et la prise de décision économique.
	Identifier les biais et les facteurs influençant la perception du public envers la politique monétaire.
	Développer des modèles d'analyse plus précis pour prédire les réactions du public et les impacts économiques des décisions de la banque centrale.

Soutenir l'inclusion et la participation du public	Accroître la transparence et la responsabilité de Bank Al-Maghrib en tenant compte des opinions et des besoins exprimés en français.
	Favoriser un dialogue inclusif et participatif sur les questions économiques et monétaires.
	Encourager la prise de conscience et l'éducation du public sur les implications de la politique monétaire pour le développement économique du Maroc.

4. Spécification des besoins

4.1. Besoins fonctionnels

- Le corpus devrait couvrir une période significative et récente afin de capturer les évolutions et les tendances actuelles de la politique monétaire. Les sources de données doivent inclure une variété de documents tels que des discours officiels, des rapports économiques, des articles de presse spécialisée et des analyses financières. Ces documents permettront d'avoir une vue d'ensemble complète et équilibrée des différentes perspectives sur la politique monétaire au Maroc.
- Analyse de sentiment : Le système doit être capable d'analyser le sentiment des textes en français en utilisant les modèles d'apprentissage automatique entraînés, en fournissant des scores de sentiment précis pour chaque document.
- Visualisation des résultats : Le système doit fournir des visualisations claires et compréhensibles des résultats de l'analyse de sentiment, permettant aux utilisateurs de comprendre rapidement les tendances et les opinions du public

4.2. Besoins non-fonctionnels

Performance : Le modèle doit être capable de traiter un grand volume de données et d'effectuer l'analyse de sentiment en temps opportun, en minimisant les temps de traitement.

Précision : Les modèles d'apprentissage automatique doivent être précis et fiables dans leur prédiction du sentiment, en assurant que les résultats reflètent fidèlement les opinions exprimées dans les textes en français.

Scalabilité : Le système doit être évolutif pour pouvoir gérer de manière efficace une augmentation du volume de données à l'avenir, sans compromettre les performances ou la précision.

5. Méthodologie de travail

Pour ce projet, nous adopterons le modèle en spirale pour guider notre approche. Le modèle en spirale est particulièrement adapté aux projets complexes comme l'apprentissage automatique, car il permet un développement itératif tout en intégrant la gestion des risques à chaque cycle. Ce modèle implique

des cycles itératifs de planification, d'analyse des risques, de développement et d'évaluation, chaque cycle étendant progressivement le précédent. En suivant le modèle en spirale, nous pouvons aborder systématiquement les incertitudes, affiner nos objectifs et nous adapter aux exigences changeantes tout au long du cycle de vie du projet. Cette approche itérative nous permet d'améliorer continuellement la qualité et l'efficacité de nos solutions d'apprentissage automatique tout en gérant efficacement les risques et les incertitudes.

6. Conclusion

Dans ce chapitre, nous avons examiné de près le contexte général entourant notre projet d'analyse des retombées médiatiques sur la politique monétaire du Maroc. À travers l'identification des problématiques et des défis rencontrés, nous avons proposé des solutions pertinentes, mettant en lumière l'importance d'un corpus de données en langue française et l'utilisation d'outils d'analyse de sentiment. En spécifiant les besoins et en détaillant notre méthodologie de travail, nous avons établi un cadre solide pour la poursuite de notre recherche. Cette exploration préliminaire nous a permis de jeter les bases nécessaires pour la prochaine étape de notre projet, à savoir la recherche bibliographique sur les outils et technologies pertinentes dans le domaine de l'analyse de données et du machine Learning.

Chapitre III : Recherche bibliographique

1. Introduction

Dans cette section, nous passerons par plusieurs technologies et outils que nous avons utilisés tout au long de notre projet. Ces outils ont été essentiels pour la collecte, le prétraitement et l'analyse des données textuelles en français, ainsi que pour le développement de modèles d'apprentissage automatique et les méthodes de visualisation de données, facilitant ainsi l'interprétation des résultats de l'analyse des sentiments.

2. Contexte du l'art

2.1. Traitement automatique des langages naturels

2.1.1. Définition du NLP

Le NLP ou le traitement automatique des langages naturels (TALN), est une branche de l'intelligence artificielle qui se concentre sur l'interaction entre les ordinateurs et les humains à travers le langage naturel. Il permet aux machines de comprendre, interpréter et générer le langage humain de manière utile tel qu'il est écrit ou parlé. Dans le contexte de la banque centrale, le NLP peut être utilisé pour analyser de grandes quantités de textes provenant de diverses sources, telles que des rapports financiers, des nouvelles économiques et des publications sur les réseaux sociaux. En utilisant le NLP, la banque centrale peut extraire des informations pertinentes, détecter les tendances et les opinions du public, et surveiller les réactions du marché en temps réel. Cela permet une prise de décision plus rapide et plus précise, en tenant compte des sentiments et des perceptions des acteurs économiques.



Figure 7: Natural language processing

2.1.2. Utilité du NLP

Le NLP joue un rôle crucial dans de nombreuses applications modernes, facilitant l'interaction entre les humains et les machines, automatisant l'analyse de grandes quantités de texte et améliorant les services de communication et de compréhension.

- **Analyse de Sentiments** : Analyse des avis clients, surveillance des médias sociaux, feedback des utilisateurs.
- **Extraction d'Informations** : Extraction des entités nommées (NER), analyse de textes juridiques, résumé automatique.
- **Traduction Automatique** : Google Translate, services de traduction en temps réel.

- **Résumé Automatique** : Résumé d'articles de recherche, de documents d'entreprise, de nouvelles.
- **Chatbots et Assistants Virtuels** : Assistants comme Siri, Alexa, et les services clients automatisés.
- **Recherche d'Informations** : Moteurs de recherche comme Google, systèmes de recommandation.
- **Classification de Textes** : Filtrage des spams, classification des e-mails, catégorisation des articles de presse.
- **Correction Automatique** : Correcteurs orthographiques comme Grammarly, outils de vérification grammaticale.
- **Génération de Textes** : Génération de contenu, rédaction automatique d'articles, création de descriptions de produits.
- **Analyse des Réseaux Sociaux** : Surveillance de la marque, analyse de la réputation, étude des tendances.

2.1.3. Le workflow du NLP

Le traitement du langage naturel (NLP) est un processus structuré qui permet de transformer des données textuelles brutes en informations exploitables. Voici les étapes typiques du workflow du NLP :

- **Collecte des Données** : Rassembler des données textuelles provenant de diverses sources telles que des articles, des blogs, des réseaux sociaux, des documents, etc. En utilisant soit le web-scraping ou APIs ou bases de données.
- **Prétraitement des Données** : Nettoyer et préparer les données pour l'analyse.
- **Extraction des Caractéristiques** : Transformer le texte en représentations numériques exploitables par des modèles de Machine Learning.
- **Modélisation** : Construire des modèles de Machine Learning pour des tâches spécifiques du NLP.
- **Évaluation** : Mesurer la performance des modèles.
- **Déploiement** : Intégrer les modèles dans des applications ou des systèmes pour une utilisation en production.
- **Maintenance et Amélioration** : Surveiller les performances des modèles et les mettre à jour régulièrement pour améliorer la précision et la fiabilité. Soit par réentraînement des modèles avec de nouvelles données ou des améliorations continues basées sur les retours des utilisateurs et les avancées technologiques.

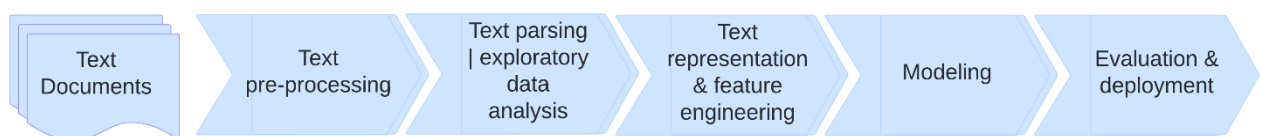


Figure 8: workflow du NLP

2.2. L'analyse des sentiments

L'analyse des sentiments [7] , aussi connue sous le nom d'analyse d'opinion, est une technique de traitement du langage naturel (NLP) utilisée pour identifier et extraire les sentiments exprimés dans un texte. Cela peut inclure des opinions, des émotions et des attitudes exprimées dans des articles de presse, des avis clients, des publications sur les réseaux sociaux, etc. Les sentiments peuvent être classifiés comme positifs, négatifs ou neutres, voire plus finement catégorisés selon une gamme d'émotions.

L'objectif principal de l'analyse de sentiment est de fournir des informations sur les opinions et les attitudes du public ou des consommateurs concernant un produit, un service, un événement ou une organisation. Pour les banques centrales, l'analyse de sentiment peut aider à comprendre les perceptions publiques et les réactions aux politiques économiques et monétaires, ainsi qu'à évaluer la confiance des consommateurs et des investisseurs. L'analyse de sentiment peut être réalisée à travers plusieurs étapes et techniques :

Tableau 4: les étapes de l'analyse des sentiments

Collecte de données	La première étape consiste à collecter les données textuelles à analyser. Cela peut inclure des tweets, des commentaires sur les réseaux sociaux, des avis clients, des articles de presse, etc.
Prétraitement des données	<ul style="list-style-type: none"> - Nettoyage des données : Eliminer les éléments non pertinents comme les balises HTML, les URLs, les ponctuations excessives et les espaces blancs. - Tokenisation : Diviser le texte en unités plus petites, telles que des mots ou des phrases. - Lemmatisation/Stemming : Réduire les mots à leur forme de base ou racine.
Extraction de caractéristiques	<p>Bag of Words (BoW) : Créer une matrice de fréquence des mots présents dans le texte.</p> <ol style="list-style-type: none"> Term Frequency – Inverse Document Frequency (TF – IDF): <ul style="list-style-type: none"> - TF mesure la fréquence d'un terme dans un document. - IDF mesure l'importance d'un terme dans l'ensemble du corpus. - La combinaison TF-IDF donne l'importance relative d'un terme dans un corpus. Word Embedding (vecteurs de texte) : <ul style="list-style-type: none"> - Représente les mots sous forme de vecteurs en basse dimension tout en préservant la similarité contextuelle. - Modèles populaires : Word2Vec et GloVe.

	<ul style="list-style-type: none"> - Utilisés dans des modèles d'apprentissage profond comme les réseaux de neurones convolutionnels (CNN) et récurrents (RNN).
Classification	Modèles de Machine Learning : Entraîner des modèles supervisés comme le Naïve Bayes, les machines à vecteurs de support (SVM), les forêts aléatoires, ou les réseaux de neurones pour classer les sentiments.
Analyse et interprétation	Interpréter les résultats pour comprendre les sentiments dominants et leur distribution.
Visualisation	Utiliser des graphiques et des visualisations pour représenter les sentiments et les tendances de manière compréhensible.

2.3. Techniques de NLP

2.3.1. Tokenisation

La tokenisation est une étape fondamentale dans tout pipeline de traitement du langage naturel (NLP). Elle consiste à découper des données textuelles non structurées en éléments discrets appelés tokens, qui peuvent être des mots, des phrases ou des sous-mots. Cette transformation est cruciale car elle convertit un texte brut en une structure de données numérique adaptée aux algorithmes d'apprentissage automatique. En fragmentant le texte en tokens, ces tokens peuvent être directement utilisés pour déclencher des actions ou des réponses automatisées, ou encore comme caractéristiques dans un pipeline de machine Learning pour prendre des décisions plus complexes.

2.3.2. Suppression du bruit

La suppression du bruit (noise removal) est une étape cruciale dans le traitement du langage naturel (NLP). Toute portion de texte qui n'est pas pertinente pour le contexte des données et le résultat final peut être considérée comme du bruit.

Exemples de bruit :

- Mots vides : Mots couramment utilisés dans une langue, comme "est", "suis", "le", "de", "dans", etc.
- URLs ou liens : Adresses web qui n'apportent pas de valeur contextuelle.
- Entités des réseaux sociaux : Mentions, hashtags, etc.
- Ponctuation : Points, virgules, points d'interrogation, etc.
- Mots spécifiques à une industrie : Terminologie propre à un secteur qui peut ne pas être pertinente pour l'analyse générale.

Méthodes pour la suppression du bruit

- Préparation d'un dictionnaire de bruit : Constituer un dictionnaire des entités bruyantes à partir

de l'ensemble de données.

- **Itération par tokens** : Parcourir le texte par tokens (ou par mots), en éliminant les tokens présents dans le dictionnaire de bruit.
- **Gestion des motifs spéciaux de bruit** : Utiliser des expressions régulières pour identifier et supprimer des motifs spécifiques de bruit, comme les URLs ou les mentions.

Un autre type de bruit textuel concerne les multiples représentations d'un même mot. Par exemple, "croissance", "croissant", "croître", "croissante" et "croissances" sont des variations du mot "croissance". Bien que leurs significations puissent différer légèrement, elles sont contextuellement similaires. Cette étape convertit toutes les disparités d'un mot en leur forme normalisée (également appelée lemma). La normalisation est une étape essentielle de l'ingénierie des caractéristiques textuelles, car elle réduit les caractéristiques à haute dimension (N caractéristiques différentes) à un espace de basse dimension (1 caractéristique), ce qui est idéal pour tout modèle d'apprentissage automatique.

Les pratiques les plus courantes de normalisation lexicale sont : la lemmatisation et le stemming, que nous expliquerons plus tard dans cette section.

2.3.3. Les stopwords

Les "stopwords" sont des mots courants dans une langue qui apparaissent très fréquemment dans les textes, mais qui ont une valeur sémantique limitée. Par exemple, en anglais, des mots comme "the", "is", "in", "and" sont considérés comme des stopwords. En français, ce seraient des mots comme "le", "la", "et", "de". L'utilisation de stopwords dans le traitement du langage naturel (NLP) permet de réduire le bruit et de se concentrer sur les mots qui portent plus de sens. Voici pourquoi ils sont importants :

- **Réduction de la dimensionnalité** : En éliminant les stopwords, on peut réduire la taille des données textuelles, ce qui rend le traitement plus rapide et efficace.
- **Amélioration de la qualité des données** : La suppression des stopwords aide à se concentrer sur les mots plus significatifs, améliorant ainsi la qualité des analyses.
- **Pré-traitement standard** : La suppression des stopwords est une étape courante dans le pré-traitement des données textuelles avant de les utiliser dans des modèles de machine Learning.

2.3.4. Lemmatisation

La lemmatisation est un processus linguistique qui consiste à réduire les mots à leur forme de base ou racine, connue sous le nom de lemme. L'objectif est de normaliser les différentes formes infléchies d'un mot afin qu'elles puissent être analysées ou comparées plus facilement. Cela est particulièrement utile en traitement du langage naturel (NLP) et en analyse de texte.

Voici comment fonctionne généralement la lemmatisation :

1. **Tokenisation** : La première étape consiste à décomposer un texte en mots ou tokens individuels. Cela peut être fait en utilisant diverses méthodes, comme la division du texte en

fonction des espaces.

2. **Étiquetage des parties du discours (POS Tagging)** : L'étiquetage des parties du discours implique d'assigner une catégorie grammaticale (comme nom, verbe, adjectif, etc.) à chaque token. La lemmatisation s'appuie souvent sur cette information, car la forme de base d'un mot peut dépendre de son rôle grammatical dans une phrase.
3. **Lemmatisation** : Une fois que chaque mot a été tokenisé et assigné à une catégorie grammaticale, l'algorithme de lemmatisation utilise un lexique ou des règles linguistiques pour déterminer le lemme de chaque mot. Le lemme est la forme de base du mot, qui peut ne pas être nécessairement la même que la racine du mot. Par exemple, le lemme de « économique » est « économie », et le lemme de « croissante » est « croître ».
4. **Application des règles** : Les algorithmes de lemmatisation s'appuient souvent sur des règles et des modèles linguistiques. Pour les verbes irréguliers ou les mots ayant plusieurs lemmes possibles, ces règles aident à prendre la bonne décision de lemmatisation.
5. **Résultat** : Le résultat de la lemmatisation est un ensemble de mots dans leur forme de base ou forme dictionnaire, ce qui facilite l'analyse et la compréhension du sens sous-jacent d'un texte.

2.3.5. Stemming

La racinisation (stemming) est le processus de réduction des mots infectés à leur racine. Par exemple, la racinisation remplace les mots " courant" par " cour". De même, les mots "finalement" et "final" sont transformés.

La racinisation consiste à enlever les derniers caractères d'un mot donné pour obtenir une forme plus courte, même si cette forme n'a pas de sens en apprentissage automatique.

Dans les cas d'utilisation du NLP tels que l'analyse des sentiments, obtenir le mot de base est important pour savoir si le mot est positif ou négatif. La racinisation est utilisée pour obtenir ce mot de base. Par exemple, dans une analyse de sentiment, il est crucial de savoir que « Croissance », « Croître » et « Croissant » proviennent tous de la racine « Croître » pour comprendre que ces mots véhiculent une connotation positive.

2.3.6. Reconnaissance d'entités nommées (NER)

La reconnaissance des entités nommées (NER) est une technique clé du traitement du langage naturel (NLP) visant à identifier et à classer les entités dans un texte non structuré. NER extrait automatiquement des informations structurées, permettant aux machines de comprendre et de catégoriser des entités telles que les noms de personnes, les organisations, les lieux, et autres expressions importantes.

Fonctionnement de la NER

1. **Détection des Entités** : Le système NER analyse le texte pour identifier et localiser les entités nommées.

2. Classification des Entités : Les entités sont classées dans des catégories prédéfinies comme personnes, organisations, lieux, etc.

Méthodes de NER

- Basée sur le Lexique : Utilisation d'un dictionnaire de termes. Cette méthode nécessite une mise à jour constante du dictionnaire.
- Basée sur les Règles : Application de règles prédéfinies basées sur des motifs et le contexte.
- Basée sur le Machine Learning : Utilisation d'algorithmes de machine Learning et de Deep Learning.

2.3.7. Feature Engineering

L'ingénierie des caractéristiques [8] sur les données textuelles est essentielle pour transformer des données non structurées en un format utilisable par les modèles d'apprentissage automatique.

Voici un résumé des principales techniques :

N-grams : Les n-grams sont des séquences de n éléments (mots, caractères, etc.) extraits d'un texte. Ils sont utilisés pour capturer les relations locales entre les mots ou les caractères dans un texte. Par exemple

- Unigram (1-gram) : "taux", "d'inflation", "élevé"
- Bigram (2-gram) : "taux d'inflation", "inflation élevé"
- Trigram (3-gram) : "taux d'inflation élevé"

L'utilisation des n-grams permet de :

- Capturer le contexte : Les n-grams aident à capturer le contexte des mots, améliorant ainsi la compréhension du texte par les modèles de machine Learning.
- Analyse de la co-occurrence : Ils permettent d'analyser comment les mots se groupent ensemble dans un texte, ce qui est utile pour des tâches comme la génération de texte et l'analyse de la syntaxe.
- Détection de motifs : Les n-grams aident à identifier des motifs fréquents dans le texte, utiles pour les modèles de prédiction de texte et la reconnaissance d'entités nommées.

TF-IDF (Term Frequency-Inverse Document Frequency) : est une méthode de pondération utilisée pour évaluer l'importance d'un mot dans un document par rapport à une collection de documents (corpus). La méthode combine deux mesures :

- **Term Frequency (TF)** : Mesure de la fréquence d'un terme dans un document donné. Plus un terme apparaît souvent dans un document, plus son score TF est élevé.
- **Inverse Document Frequency (IDF)** : Mesure de la rareté d'un terme dans l'ensemble des documents. Plus un terme est rare, plus son score IDF est élevé.

La formule du TF-IDF est donnée par :

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t)$$

Où t est le terme, d est le document, et $IDF(t) = \log(N / nt)$ avec N le nombre total de documents et nt le nombre de documents contenant le terme t .

Les avantages du TF-IDF incluent :

- **Pondération** : Le TF-IDF accorde plus de poids aux mots rares et pertinents, ce qui améliore la qualité de la représentation des documents.
- **Filtrage du bruit** : En réduisant l'importance des mots fréquents (stopwords), le TF-IDF aide à filtrer le bruit et à se concentrer sur les termes significatifs.
- **Applications diverses** : Utilisé dans des tâches telles que la recherche d'information, la classification de texte et le clustering.

Count Vectorizer : Le Count Vectorizer est une technique de conversion de texte en données numériques, en comptant le nombre d'occurrences de chaque mot dans un document. Chaque document est représenté par un vecteur de longueur fixe, où chaque dimension représente un mot unique du corpus, et la valeur dans chaque dimension correspond au nombre d'occurrences de ce mot dans le document.

Avantages du Count Vectorizer :

- **Simplicité** : Facile à comprendre et à implémenter, il constitue une base pour de nombreuses tâches de NLP.
- **Représentation des documents** : Transforme les documents textuels en vecteurs numériques, rendant possible leur utilisation dans les algorithmes de machine Learning.

Préparation des données : Sert de première étape dans le prétraitement des données textuelles avant d'appliquer des techniques plus avancées comme le TF-IDF ou les modèles de Word embedding.

2.4. Les modèles de machine learning

2.4.1. Naive bayes

Naive Bayes [9] est une famille de classificateurs probabilistes simples basés sur le théorème de Bayes, avec une forte hypothèse d'indépendance entre les caractéristiques. En d'autres termes, Naive Bayes suppose que la présence ou l'absence d'une caractéristique particulière dans une classe donnée est indépendante des autres caractéristiques. Malgré cette hypothèse simplifiée, Naive Bayes a prouvé son efficacité dans de nombreux domaines, notamment le filtrage de spam, la classification de texte, et la reconnaissance de formes.

Le théorème de Bayes

Le théorème de Bayes est utilisé pour calculer la probabilité conditionnelle d'un événement. Il est exprimé mathématiquement comme suit :

$$P(A|B) = [P(B|A) \cdot P(A)] / P(B)$$

Où :

$P(A|B)$ est la probabilité de l'événement A étant donné l'événement B.

$P(B|A)$ est la probabilité de l'événement B étant donné l'événement A.

$P(A)$ est la probabilité a priori de l'événement A.

$P(B)$ est la probabilité a priori de l'événement B.

Hypothèse d'indépendance conditionnelle

Naive Bayes suppose que les caractéristiques du document sont conditionnellement indépendantes les unes des autres étant donné la classe. Cela signifie que :

$$P(d|c) = P(x_1, x_2, \dots, x_n|c) = P(x_1|c) \cdot P(x_2|c) \cdot \dots \cdot P(x_n|c)$$

Où 1, 2, ..., x1, x2, ..., xn sont les caractéristiques (par exemple, les mots dans un document).

Calcul des probabilités

Pour chaque classe c et chaque caractéristique xi dans le document, nous devons estimer deux types de probabilités :

La probabilité a priori P, qui est la proportion de documents appartenant à la classe c.

La probabilité conditionnelle $P(x_i|c)$, qui est la proportion de documents de la classe c contenant la caractéristique xi.

Classification

Pour classer un nouveau document d, nous calculons le score pour chaque classe c et nous choisissons ensuite la classe c qui maximise ce score :

$$P(c|d) \propto P \cdot \prod_{i=1}^n P(x_i|c)$$

2.4.2. SVM (Support Vector Machine)

Les Machines à Vecteurs de Support (SVM) [10] sont des algorithmes d'apprentissage supervisé puissants utilisés pour des tâches de classification et de régression en apprentissage automatique. Elles sont particulièrement efficaces dans les espaces de haute dimension et sont connues pour leur robustesse et leur précision.

L'objectif des SVM est de trouver l'hyperplan optimal qui sépare le mieux les points de données de différentes classes dans un espace de caractéristiques. L'"hyperplan" optimal est défini comme celui qui maximise la marge entre les classes. La marge est la distance entre l'hyperplan et les points de données les plus proches de chaque classe, appelés vecteurs de support.

Le Concept d'Hyperplan

Dans un espace de n dimensions, un hyperplan est un sous-espace affine plat d'une dimension inférieure à celle de son espace ambiant. Par exemple, dans un espace bidimensionnel, un hyperplan est une ligne, tandis que dans un espace tridimensionnel, c'est un plan.

La représentation mathématique d'un hyperplan est donnée par :

$$\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = 0$$

Où :

- \mathbf{w} est un vecteur de poids (coefficients).
- \mathbf{x} est un vecteur représentant un point dans l'espace.
- \mathbf{b} est le terme de biais.

Maximiser la Marge

L'objectif principal des SVM est de trouver l'hyperplan qui maximise la marge entre les classes. La marge est définie comme suit :

$$\text{Marge} = 2 / \|\mathbf{w}\|$$

Pour maximiser la marge, les SVM doivent minimiser $\|\mathbf{w}\|$, sous la contrainte que tous les points sont correctement classifiés. Cela conduit au problème d'optimisation suivant :

$$\text{Min}_{\mathbf{w}, \mathbf{b}} \frac{1}{2} \|\mathbf{w}\|^2$$

$$\text{Soumis à } y_i (\mathbf{w} \cdot \mathbf{x}_i + \mathbf{b}) \geq 1$$

Où y_i est l'étiquette de classe du point \mathbf{x}_i , qui peut être +1 ou -1.

Multiplicateurs de Lagrange et Problème Dual

Pour résoudre le problème d'optimisation, les SVM utilisent des multiplicateurs de Lagrange, le transformant en un problème dual. La forme duale du problème est plus traitable, surtout pour les grands ensembles de données. Elle est donnée par :

$$\text{Max}_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j)$$

$$\text{Soumis à } \sum_i \alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq C$$

Où α_i sont les multiplicateurs de Lagrange, et « C » est un paramètre de régularisation qui contrôle le compromis entre la maximisation de la marge et la minimisation des erreurs de classification.

Le Trick du Noyau

Dans les scénarios où les données ne sont pas linéairement séparables, les SVM utilisent le trick du noyau pour mapper les données dans un espace de plus haute dimension où un hyperplan linéaire peut

être utilisé pour séparer les données. Cela se fait à l'aide de fonctions de noyau, qui calculent le produit scalaire des points de données dans l'espace transformé sans effectuer explicitement la transformation. Les fonctions de noyau courantes incluent :

- **Noyau Linéaire** : $x_i \cdot x_j$
- **Noyau Polynomial** : $(x_i \cdot x_j + 1)^d$
- **Noyau à Fonction de Base Radiale (RBF)** : $\exp(-\gamma \|x_i - x_j\|^2)$

3. Outils et logiciels

3.1. Excel

Excel a été utilisé comme un outil polyvalent pour la gestion et la manipulation de données. Il a permis d'effectuer des tâches telles que l'organisation des données brutes, le filtrage, le tri et la création de tableaux de bord simples pour une visualisation rapide.

3.2. Visual Studio Code (VSC)

Visual Studio Code, souvent abrégé en VSC, est un éditeur de code source développé par Microsoft. Il est léger mais puissant, et supporte une large gamme de langages de programmation grâce à ses nombreuses extensions.

3.3. PyCharm

PyCharm est un environnement de développement intégré (IDE) développé par JetBrains spécifiquement pour le développement en Python. Il est largement utilisé par les développeurs Python en raison de ses fonctionnalités avancées et de son support complet pour le langage.

3.4. Lucidchart

Lucidchart est un outil de création de diagrammes basé sur le cloud, utilisé pour visualiser des processus, des systèmes et des structures de données. Il est largement utilisé par les équipes de développement, les analystes business et les ingénieurs pour planifier et documenter des projets.

3.5. Google Colab

Google Colab a été choisi comme environnement de développement pour exécuter du code Python dans le cloud. Grâce à ses fonctionnalités de partage et d'exécution de code en ligne, il a facilité la collaboration et l'accès aux ressources de calcul, tout en offrant une compatibilité native avec les bibliothèques Python couramment utilisées.

4. Frameworks et langages de programmations

4.1. Node.js

Node.js est un environnement d'exécution JavaScript côté serveur basé sur le moteur JavaScript V8 de Google Chrome. Il permet aux développeurs de créer des applications serveur rapides et évolutives en utilisant JavaScript.

4.2. Next.js

Next.js [11] est un Framework React pour le développement d'applications web côté serveur et côté client. Développé par Vercel, il est connu pour ses fonctionnalités avancées et sa flexibilité.

4.3. Python

Python a servi de langage de programmation principal pour le développement de ce projet. Sa simplicité, sa polyvalence et sa riche bibliothèque de modules en ont fait un choix idéal pour le traitement des données, l'analyse statistique et la mise en œuvre d'algorithmes d'apprentissage automatique.

4.4. Tailwind CSS

Tailwind CSS est un framework CSS unique en son genre qui fournit des classes utilitaires configurables à la place des styles CSS traditionnels. Cela permet aux développeurs de composer rapidement des interfaces utilisateur responsives en HTML sans avoir à écrire de CSS personnalisé volumineux.

4.5. Bibliothèques Python

Les bibliothèques qui suivent couvrent un large éventail de fonctionnalités, allant de la manipulation de texte et des données, à la visualisation graphique, en passant par le traitement du langage naturel et l'apprentissage automatique.

Tableau 5: bibliothèques de python

Traitement du texte et du langage naturel	Re	Permet la manipulation et la recherche de motifs dans des chaînes de caractères en utilisant des expressions régulières.
	Spacy	Est une bibliothèque de traitement du langage naturel (NLP) rapide et avancée, utilisée pour des tâches comme le tokenisation, l'analyse syntaxique, et la reconnaissance d'entités nommées.
	Nltk	- word_tokenize, sent_tokenize : Fonctions de tokenisation de mots et de phrases.

		<ul style="list-style-type: none"> - PorterStemmer : Algorithme de racinisation pour réduire les mots à leur racine. - stopwords : Module pour manipuler les mots vides (mots courants à ignorer). - SentimentIntensityAnalyzer : Outil pour l'analyse de sentiments.
	Unidecode	La bibliothèque `unidecode` convertit les caractères Unicode en leur équivalent ASCII le plus proche, utile pour normaliser les textes.
	Textstat	Fournit des outils pour évaluer la complexité et la lisibilité des textes à l'aide de divers indices et formules.
Analyse et manipulation des données	Numpy	Est une bibliothèque essentielle pour le calcul scientifique en Python, offrant un support pour les tableaux multidimensionnels et des fonctions mathématiques de haut niveau.
	Collections	Le module `collections` fournit des structures de données alternatives comme `Counter`, qui est utilisé pour compter les occurrences d'éléments dans une séquence.
Visualisation des données	Matplotlib	Est une bibliothèque de visualisation de données qui permet de créer des graphiques en 2D, comme des histogrammes, des barres, des courbes et plus encore.
	Matplotlib.pyplot	Est un sous-module de `matplotlib` qui fournit une interface de tracé semblable à celle de MATLAB pour créer des graphiques.
	Wordcloud	Est une bibliothèque pour la génération de nuages de mots à partir de textes, où la taille de chaque mot est proportionnelle à sa fréquence ou à son importance.
Développement web	Flask	<ul style="list-style-type: none"> - Flask [12] : Un micro-Framework léger pour le développement d'applications web en Python. - request : Permet de gérer les données envoyées par les utilisateurs via les requêtes HTTP. - jsonify : Convertit les objets Python en JSON pour les réponses HTTP. - send_file : Permet d'envoyer des fichiers depuis le serveur vers le client.
	Flask_cors	Est une extension de Flask qui permet de gérer les requêtes Cross-Origin Resource Sharing (CORS), nécessaires pour autoriser les échanges de ressources entre différentes origines.
Sérialisation et désérialisation	Pickle	La bibliothèque `pickle` est utilisée pour la sérialisation et la désérialisation des objets Python, permettant de sauvegarder et de charger des modèles de machine Learning.
Machine learning	Sklearn (scikit-learn)	<ul style="list-style-type: none"> - CountVectorizer : Convertit une collection de documents en une matrice de comptage de tokens. - TfidfVectorizer : Convertit une collection de documents en une matrice TF-IDF.

Traitement et manipulation des images	PIL (Image)	La bibliothèque Python Imaging Library (PIL) permet de créer, manipuler et enregistrer des images dans divers formats. `Image` est un module pour ouvrir et manipuler des images.
---------------------------------------	------------------------	---

5. Conclusion

En conclusion de ce chapitre, nous avons exploré un éventail de technologies, d'outils et de techniques essentiels utilisés dans ce projet. De l'ingénierie des caractéristiques à l'analyse des sentiments, en passant par le prétraitement des données et l'entraînement des modèles, chaque étape a été soigneusement sélectionnée pour répondre aux exigences spécifiques de notre analyse. Ces fondations techniques sont cruciales pour la conception et l'exécution réussie du système que nous allons détailler dans le prochain chapitre. Nous aborderons en détail la mise en œuvre pratique de ces techniques sur nos données, ainsi que les résultats obtenus lors de l'entraînement de nos modèles. Cette phase clé du projet vise à transformer les données brutes en informations exploitables, tout en fournissant un cadre robuste pour une analyse approfondie dans les chapitres à venir.

Chapitre IV : Ingénierie et analyse conceptuelle

1. Introduction

Dans ce chapitre, nous allons explorer le côté de conceptuel de notre projet. Nous commencerons par détailler le diagramme en cas d'utilisation du système et le diagramme d'activité qui illustre le flux de travail et les processus clés impliqués. Ensuite, nous aborderons la collecte et le prétraitement des données, en décrivant les sources et les méthodes utilisées pour obtenir les articles de presse pertinents. Enfin, nous expliquerons en profondeur l'entraînement des modèles et les résultats obtenus.

2. Conception

Pour concevoir et structurer notre projet, nous avons utilisé la méthode de conception UML (Unified Modeling Language). UML est une norme de modélisation largement utilisée en ingénierie logicielle pour représenter graphiquement les différents aspects d'un système, allant des structures statiques aux comportements dynamiques. Cette méthode nous permet de visualiser, spécifier, construire et documenter les composants de notre projet de manière claire et systématique.

2.1. Diagramme en cas d'utilisation

Un diagramme de cas d'utilisation est un outil visuel qui représente les interactions entre les utilisateurs (ou acteurs) et le système pour montrer comment ce dernier répond à leurs besoins. Dans ce contexte, le diagramme en cas d'utilisation du système de tableau de bord pour l'analyse des sentiments des articles décrit les différentes fonctionnalités offertes par le système et comment les utilisateurs interagissent avec elles.

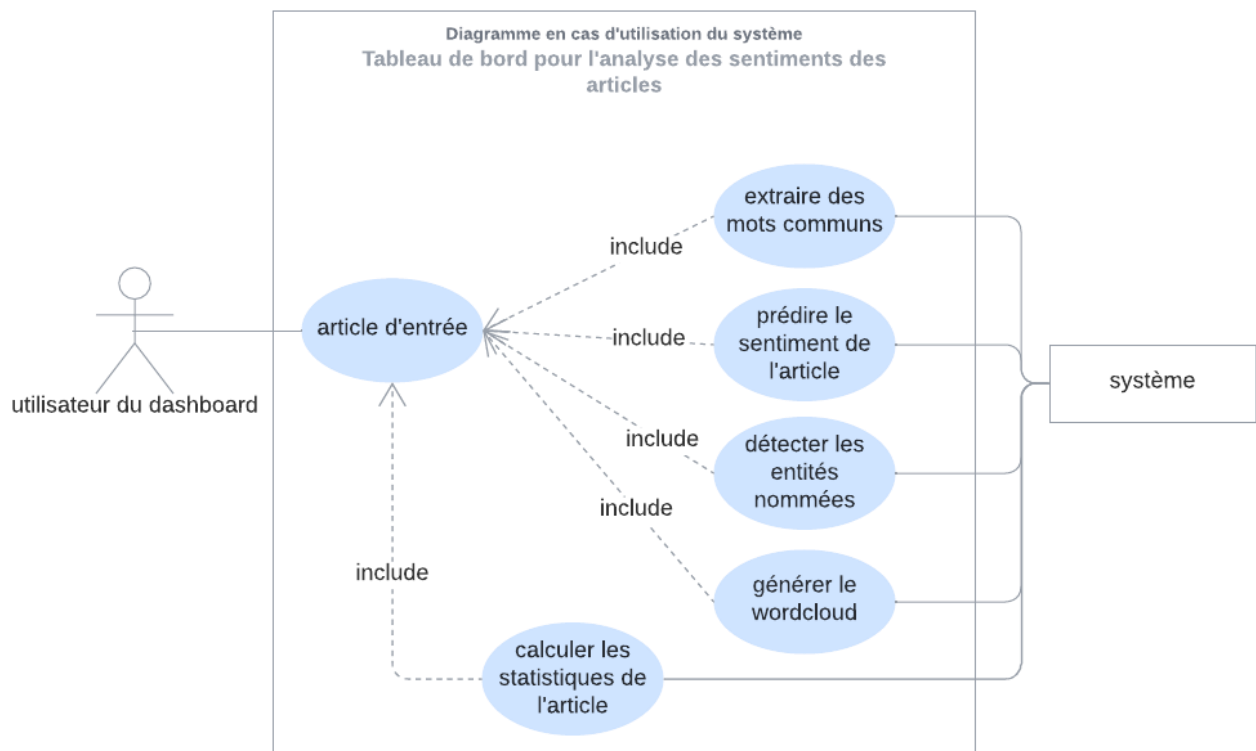


Figure 9: diagramme en cas d'utilisation

Les acteurs peuvent entrer un article, ce qui inclut plusieurs processus : générer un nuage de mots, extraire des mots communs, prédire le sentiment de l'article, détecter les entités nommées et calculer les statistiques de l'article. Chaque fonction principale est représentée par une flèche "include" indiquant qu'elle fait partie du processus global d'analyse des sentiments. Par exemple, lorsqu'un utilisateur entre un article, le système génère un nuage de mots, extrait les mots communs, prédit le sentiment de l'article, détecte les entités nommées et calcule les statistiques de l'article.

2.2. Diagramme d'activité

Pour optimiser notre modèle de manière efficace, nous devons suivre un processus rigoureux et structuré. Tout d'abord, nous collectons les données nécessaires, en veillant à obtenir un ensemble de données suffisamment large et représentatif. Ensuite, nous inspectons minutieusement ces données pour identifier toute anomalie ou problème potentiel. Cette étape est cruciale pour garantir la qualité des données que nous allons utiliser.

Une fois les données inspectées, nous procédons au nettoyage et à la prétraitement des données. Cela inclut la gestion des valeurs manquantes, la normalisation ou la mise à l'échelle des caractéristiques, et l'encodage des variables catégorielles. Ce processus permet de préparer les données pour l'étape suivante.

Les données nettoyées sont ensuite divisées en deux ensembles : un ensemble de données d'entraînement et un ensemble de données de test. Cette séparation est essentielle pour évaluer la performance du modèle de manière objective. Nous utilisons l'ensemble de données d'entraînement pour entraîner notre modèle, en ajustant ses paramètres afin de minimiser l'erreur sur cet ensemble.

Après l'entraînement, nous testons notre modèle sur l'ensemble de données de test pour prédire avec le modèle entraîné et obtenir une mesure de sa performance. Si les métriques de performance obtenues sont acceptables, nous validons le modèle. Cependant, si les résultats ne sont pas satisfaisants, nous devons évaluer le modèle en détail pour identifier les points faibles.

Cette évaluation nous guide dans la fine-tuning du modèle, ce qui peut inclure l'ajustement des hyperparamètres, l'essai de nouvelles architectures de modèle, ou l'amélioration de la prétraitement des données. Ce processus itératif se poursuit jusqu'à ce que nous obtenions un modèle dont les performances répondent à nos critères. Ainsi, nous nous assurons que notre modèle est non seulement précis mais aussi robuste et généralisable à de nouvelles données.

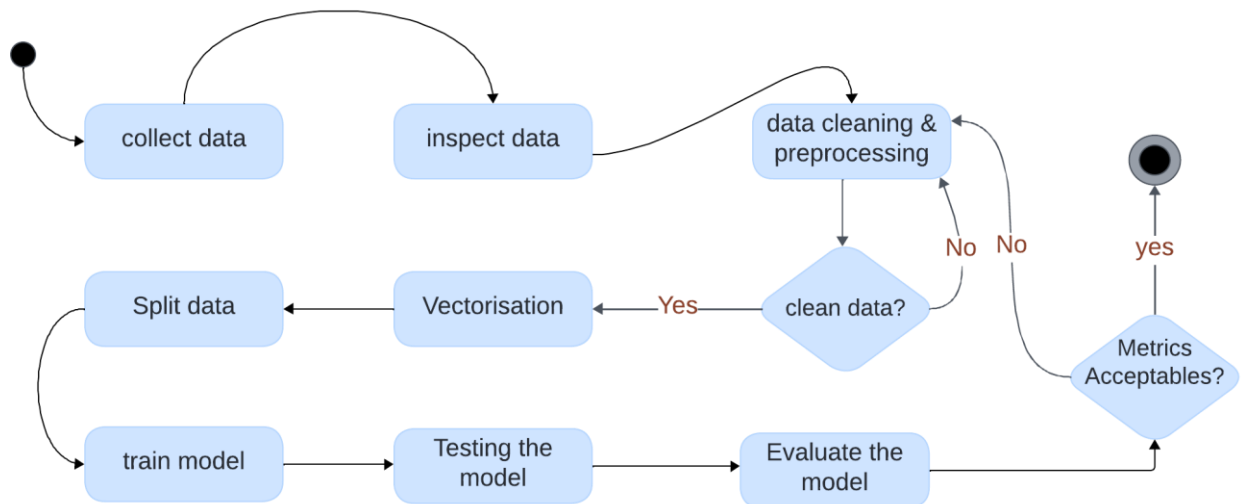


Figure 10: diagramme d'activité du processus du projet

3. Collecte des données

Les données utilisées pour ce projet ont été initialement fournies par le service Media de la Direction Autonome de Communication au sein de BAM. Après une réunion avec deux membres du personnel, où nous avons discuté de notre projet, de ses objectifs et de nos besoins, ils ont aimablement mis à notre disposition un ensemble de données contenant des articles écrits par la presse nationale et provenant de différents sites web. Ces données nous ont été envoyées par email sous forme de fichiers Excel, classés par année pour chaque fichier. Chaque article y est minutieusement décrit par son ton, sa source, sa date, son sujet et d'autres métriques détaillées et pertinentes. Pour notre projet, nous aurons simplement avoir besoin des deux colonnes contenant les articles et leur ton (cible) pour notre étude de cas.

Les données couvrent la période de 2020 à 2024. Elles sont initialement étiquetées sur une plateforme dédiée à la collecte d'articles de presse liés à la banque centrale, à ses politiques et aux différentes décisions prises, entre autres. Cette plateforme est utilisée pour saisir ces articles, les analyser manuellement et attribuer à chacun une tonalité, également manuellement. À la fin, la plateforme peut générer des graphiques et des visualisations pour illustrer l'impact d'une décision ou d'un phénomène économique, et exporter des rapports à ce sujet.

Avant de procéder au prétraitement de ces données, nous devons les inspecter et filtrer les articles qui ne sont pas liés à la politique monétaire de la banque centrale, vérifier les données manquantes ou dupliquées, et s'assurer que les articles sont bien étiquetés. Une fois cette inspection réalisée, nous pourrions passer aux étapes suivantes du prétraitement.

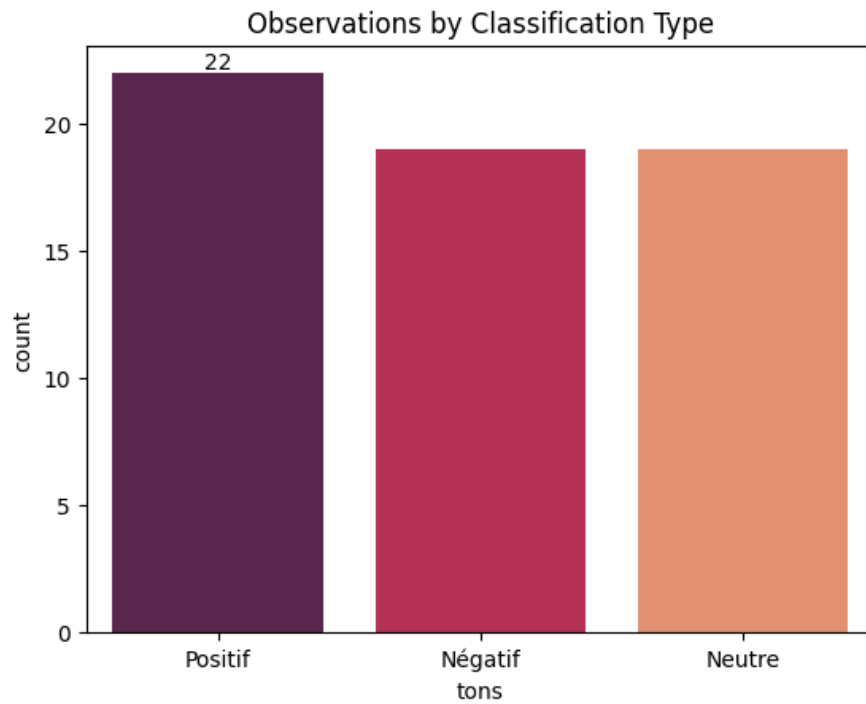


Figure 11: observation des articles par type de classification

4. Prétraitement des données

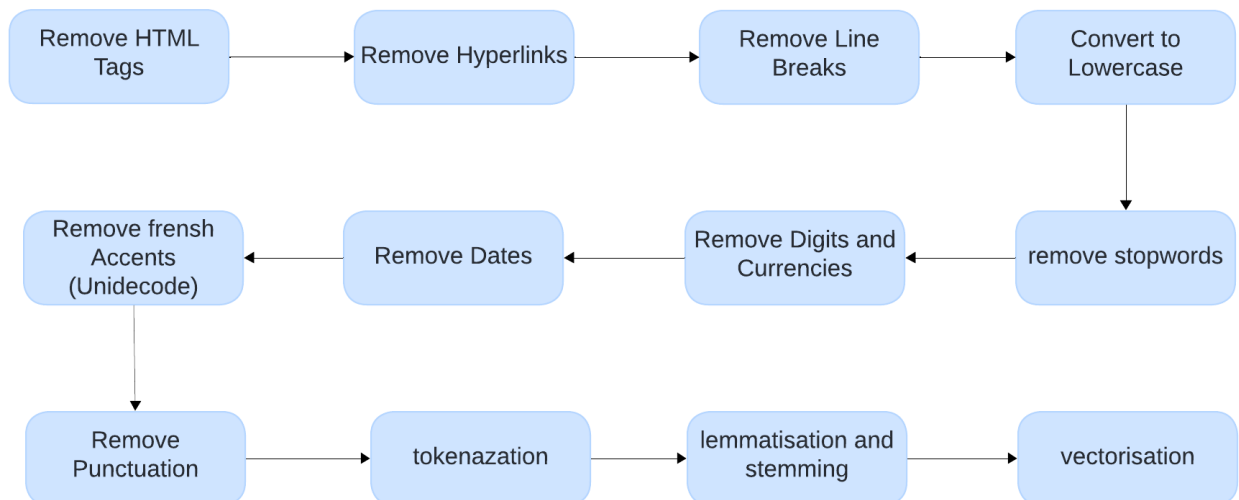


Figure 12: processus du prétraitement des données

Pour le prétraitement de nos données textuelles, nous avons suivi une série d'étapes méthodiques afin de garantir que les textes soient prêts pour l'analyse. Tout d'abord, nous avons supprimé les balises HTML pour éliminer tout code potentiellement intrusif. Ensuite, les hyperliens présents dans le texte ont été enlevés pour éviter les distractions et les biais dans l'analyse. Les sauts de ligne ont également été supprimés pour uniformiser le texte. Nous avons ensuite converti le texte en minuscules pour garantir la cohérence dans le traitement.

Une fois ces étapes de nettoyage initial réalisées, nous avons supprimé les accents français à l'aide de

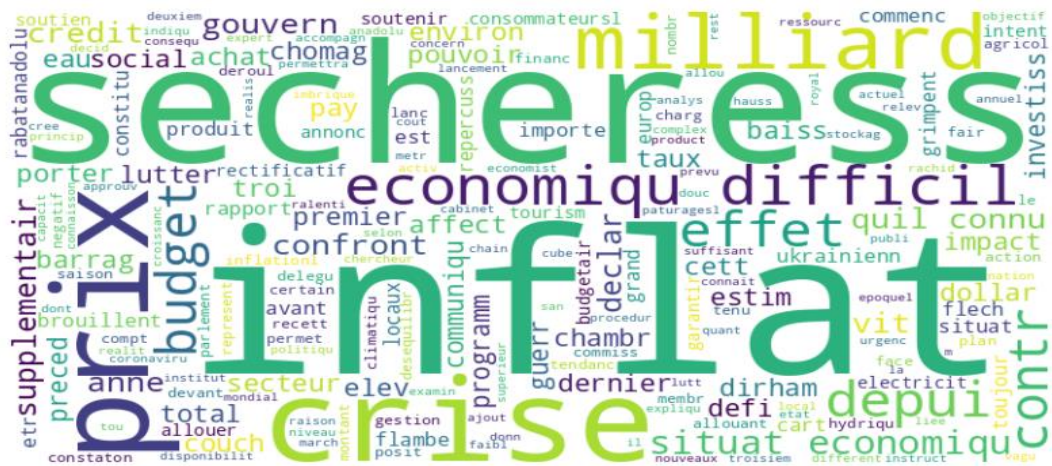


Figure 13: word-cloud des 2 articles du Corpus

6. Résultat de l'entraînement et tests du modèle

6.1. Définition des mesures de performances :

Accuracy : La précision est le rapport entre le nombre de prédictions correctes et le nombre total de prédictions. Elle mesure la proportion des échantillons correctement classés par le modèle.

Formule : $\text{Précision} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombre total de prédictions}}$

Précision : La précision est le rapport entre les vrais positifs et la somme des vrais positifs et des faux positifs. Elle mesure la capacité du modèle à ne pas étiqueter un échantillon négatif comme positif.

Formule : $\text{Précision} = \frac{\text{VraiPositif}}{(\text{VraiPositif} + \text{FauxPositif})}$

Rappel (Recall) : Le rappel est le rapport entre les vrais positifs et la somme des vrais positifs et des faux négatifs. Il mesure la capacité du modèle à identifier tous les échantillons positifs.

Formule : $\text{Rappel} = \frac{\text{VraiPositif}}{(\text{VraiPositif} + \text{FauxNégatif})}$

F1-score : Le F1-score est la moyenne harmonique de la précision et du rappel. Il prend en compte à la fois les faux positifs et les faux négatifs et est particulièrement utile lorsque on a besoin d'un équilibre entre la précision et le rappel.

Formule : $\text{F1-score} = 2 \times \left[\frac{(\text{Précision} \times \text{Rappel})}{(\text{Précision} + \text{Rappel})} \right]$

6.2. Les résultats obtenus

6.2.1. Naive Bayes

Les résultats obtenus en utilisant le Naive Bayes avec CountVectorizer et stemming et suppression des stopwords sont les suivants :

Tableau 6: metrics du Naive Bayes

Accuracy	72.22 %
Precision	75.69 %
Recall	72.22 %
F1-score	71.85 %

Accuracy : La précision du modèle Naive Bayes est de **72.22 %**. Cela signifie que le modèle a correctement classé 72.22 % des échantillons totaux. La précision mesure le rapport global entre les prédictions correctes et le nombre total de prédictions.

Précision : La précision du modèle est de **75.69 %**. Cela signifie que parmi toutes les prédictions positives faites par le modèle, 75.69 % étaient correctes. La précision évalue la capacité du modèle à éviter les faux positifs, c'est-à-dire à ne pas étiqueter incorrectement un échantillon négatif comme positif.

Recall : Le rappel du modèle est de 72.22 %. Cela indique que le modèle a correctement identifié 72.22 % des échantillons positifs réels. Le rappel mesure la capacité du modèle à détecter tous les échantillons positifs, c'est-à-dire à minimiser les faux négatifs.

F1-score : Le F1-score du modèle est de **71.85 %**. Le F1-score est la moyenne harmonique de la précision et du rappel. Il prend en compte à la fois les faux positifs et les faux négatifs, offrant ainsi une mesure globale de la performance du modèle. Un F1-score de 71.85 % indique que le modèle maintient un bon équilibre entre la précision et le rappel.

6.2.2. SVM

Les résultats obtenus en utilisant le SVM avec TfidfVectorizer et stemming et suppression des stopwords sont les suivants :

Tableau 7: Metrics du SVM

Accuracy	50.0 %
Precision	31.67 %
Recall	50.0 %
F1-score	38.56 %

Accuracy : La précision du modèle SVM est de **50.0 %**. Cela signifie que le modèle a correctement classé 50 % des échantillons totaux. La précision globale mesure le rapport entre les prédictions correctes et le nombre total de prédictions.

Précision : La précision du modèle SVM est de **31.67 %**. Cela signifie que parmi toutes les prédictions positives faites par le modèle, seulement 31.67 % étaient correctes. La précision évalue la capacité du modèle à éviter les faux positifs, c'est-à-dire à ne pas étiqueter incorrectement un échantillon négatif comme positif.

Rappel (Recall) : Le rappel du modèle SVM est de **50.0 %**. Cela indique que le modèle a correctement identifié 50 % des échantillons positifs réels. Le rappel mesure la capacité du modèle à détecter tous les échantillons positifs, c'est-à-dire à minimiser les faux négatifs.

F1-score : Le F1-score du modèle SVM est de **38.56 %**. Le F1-score est la moyenne harmonique de la précision et du rappel. Il prend en compte à la fois les faux positifs et les faux négatifs, offrant ainsi une mesure globale de la performance du modèle. Un F1-score de 38.56 % indique que le modèle a des difficultés à maintenir un bon équilibre entre la précision et le rappel.

6.2.3. Confusion matrix

Une matrice de confusion est une représentation tabulaire qui résume la performance d'un modèle de classification en présentant les comptages des prédictions vraies positives, vraies négatives, fausses positives et fausses négatives. Nous présenterons la matrice de confusion générée par nos modèles pour visualiser leur précision et leurs taux d'erreur pour différentes classes de sentiments.

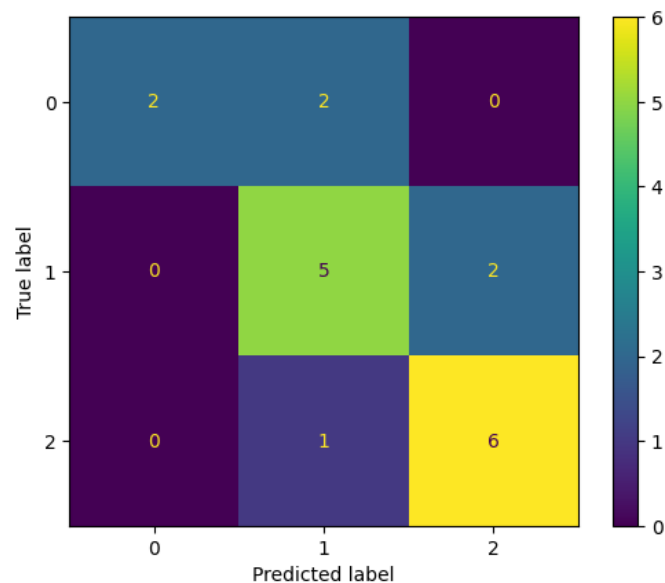


Figure 14: matrice de Confusion / Naive Bayes

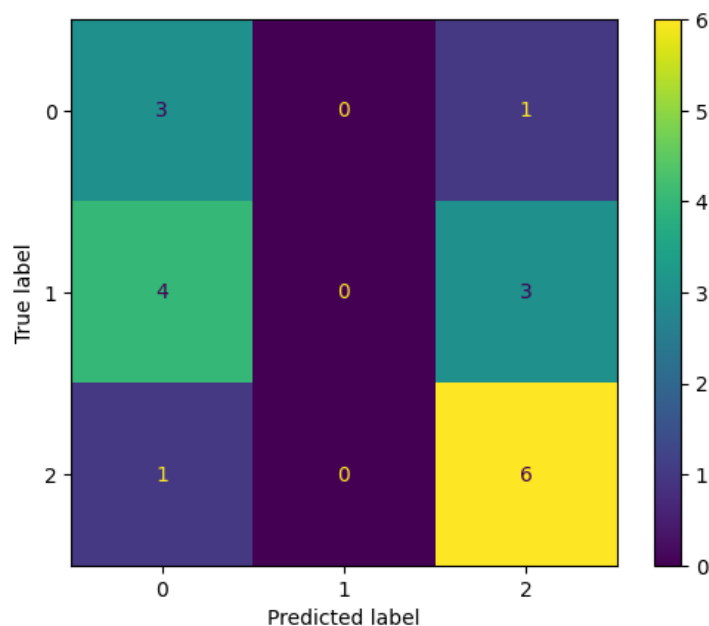


Figure 15: matrice de confusion / SVM

Le modèle Naive Bayes semble avoir obtenu de bons résultats sur cette tâche, car on observe un nombre élevé de classifications correctes le long de la diagonale de la matrice de confusion.

En revanche, le modèle SVM semble avoir obtenu des résultats moins bons que le modèle Naive Bayes sur cette tâche, car on observe un nombre plus faible de classifications correctes le long de la diagonale de la matrice de confusion.

6.3. Pourquoi Naive Bayes ?

Pour choisir le meilleur algorithme en fonction des caractéristiques de notre dataset et du matériel disponible, on doit considérer plusieurs critères. Le Naive Bayes fonctionne bien avec un grand volume de textes et des ressources matérielles limitées, grâce à sa simplicité et rapidité. La classification est adaptée aux datasets de taille moyenne à grande et offre une bonne précision avec un prétraitement adéquat, nécessitant des ressources modérées. Les SVM, bien que très précis, exigent des ressources élevées et un bon prétraitement, et conviennent mieux aux datasets de petite à moyenne taille. Pour notre corpus d'articles, le Naive Bayes apparaît comme un choix judicieux en raison de sa simplicité et de son efficacité sur les données textuelles avec des ressources limitées.

Tableau 8: Etude comparative des modèles de ML

Critère	Naive Bayes	SVM
Taille du Dataset	Fonctionne bien avec de grands datasets textuels.	Performant avec des datasets de petite à moyenne taille.
Précision et Performance	Très performant pour les grandes quantités de données textuelles, surtout avec des caractéristiques conditionnellement indépendantes.	Très précis avec un bon réglage des paramètres, mais coûteux en calcul.
Temps d'Entraînement	Rapide.	Long, surtout avec de grandes données.
Complexité Computationnelle	$O(n)$ en temps d'entraînement.	$O(n^2)$ à $O(n^3)$ en temps d'entraînement.
Interprétabilité	Facile à interpréter avec des probabilités de chaque classe.	Moins interprétable, surtout avec noyaux non linéaires.
Robustesse aux Données Bruitées	Sensible aux caractéristiques non pertinentes, mais robuste avec un bon prétraitement.	Très robuste avec des noyaux appropriés.
Facilité de Mise en Œuvre	Facile à implémenter et comprendre, peu de réglages de paramètres.	Complexe à implémenter, nécessite un bon réglage des paramètres.

7. Conclusion

Ce chapitre nous a permis de plonger dans les aspects conceptuels de notre projet. Nous avons commencé par élaborer les diagrammes de cas d'utilisation et d'activité, mettant en lumière le flux de travail et les processus clés impliqués dans notre système. Ensuite, nous avons détaillé la collecte et le prétraitement des données, en explorant les sources et les méthodes utilisées pour recueillir les articles de presse pertinents. Enfin, nous avons approfondi l'entraînement des modèles et analysé les résultats obtenus. Ces fondations solides nous préparent à présent à présenter dans le chapitre suivant

l'interface que nous avons développée, concrétisée sous forme de tableau de bord, illustrant ainsi notre progression vers une application pratique et utilisable pour notre cas d'utilisation.

Chapitre V : Réalisation

1. Introduction

Ce dernier chapitre de notre rapport se concentre sur la concrétisation de notre projet à travers une interface utilisateur intuitive et un tableau de bord fonctionnel. Nous présenterons ici comment nous avons intégré les modèles d'analyse de sentiment et les techniques de visualisation pour offrir une expérience informative et interactive. Cette interface joue un rôle crucial dans la mise en valeur des résultats de notre analyse et dans la facilitation de l'accès aux informations clés sur les décisions de politique monétaire au Maroc.

2. Les interfaces graphiques

2.1. Formulaire d'Entrée

La première interface de l'application est un formulaire d'entrée où l'utilisateur peut saisir les informations suivantes :

- **Titre de l'article** : Champ pour entrer le titre de l'article.
- **Contenu** : Zone de texte pour entrer le contenu complet de l'article.
- **Date de publication** : Champ pour entrer la date de publication de l'article.

L'utilisateur doit remplir ces champs puis envoyer la demande pour que l'analyse soit effectuée.

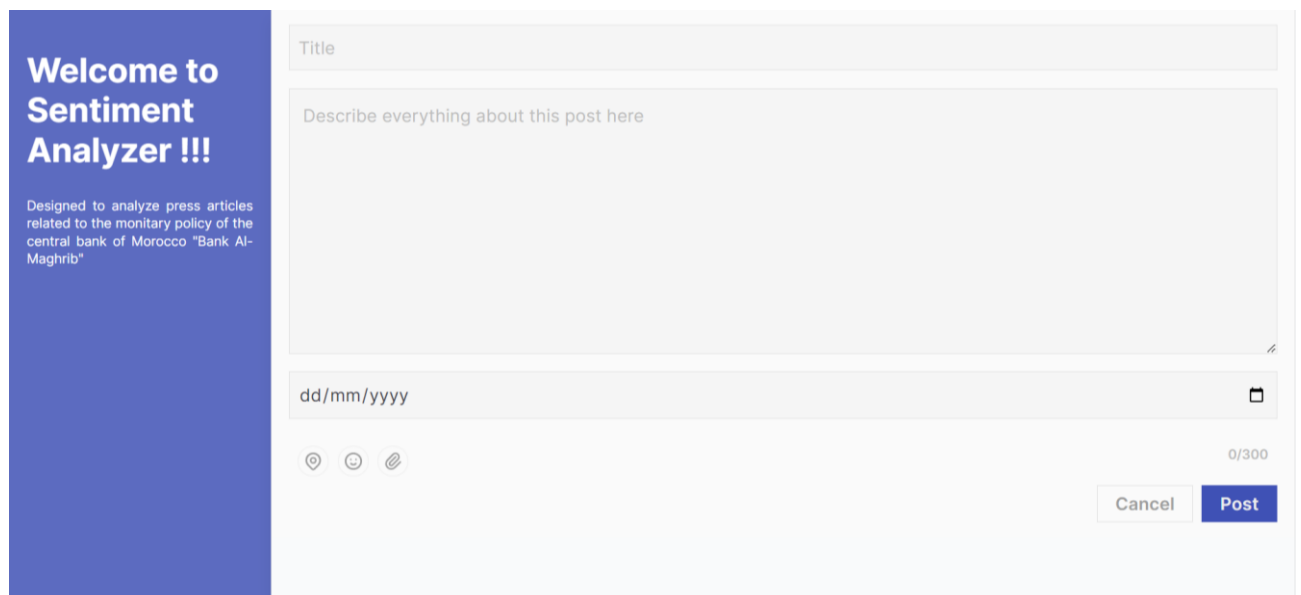


Figure 16: formulaire d'entrée

2.2. Affichage des Résultats

Une fois que l'utilisateur soumet le formulaire d'entrée, la deuxième interface s'affiche, montrant les résultats de l'analyse de l'article. Cette interface contient plusieurs éléments visuels et informatifs :

- **Graphique en Demi-Doughnut** : Ce graphique montre le sentiment le plus probable de l'article (positif, négatif ou neutre) ainsi que le pourcentage correspondant.
- **Nuage de Mots** : Représentation visuelle des mots les plus fréquents dans le texte de l'article. Plus un mot est fréquent, plus il apparaît grand dans le nuage.

- **Résumé** : Une section contenant des statistiques sur le texte d'entrée, telles que : Nombre de mots, nombre de phrases, Longueur moyenne des phrases et score de lisibilité
- **Graphique à Barres** : Ce graphique montre les 5 mots les plus courants dans le texte de l'article, fournissant une vue rapide des termes clés utilisés.
- **Reconnaissance des Entités Nommées (NER)** : Cette section identifie et classe les entités nommées présentes dans le texte, telles que les noms de personnes, d'organisations, de lieux, etc.
- **Analyse des sentiments par sections** : Ce type de tableau de bord permet d'évaluer la tonalité générale d'un texte en fonction de la couleur attribuée aux différents segments. Dans ce cas précis :
 - Rouge indique une tonalité négative du texte analysé.
 - Vert indique une tonalité positive du texte analysé.
 - Bleu indique une tonalité neutre du texte analysé.

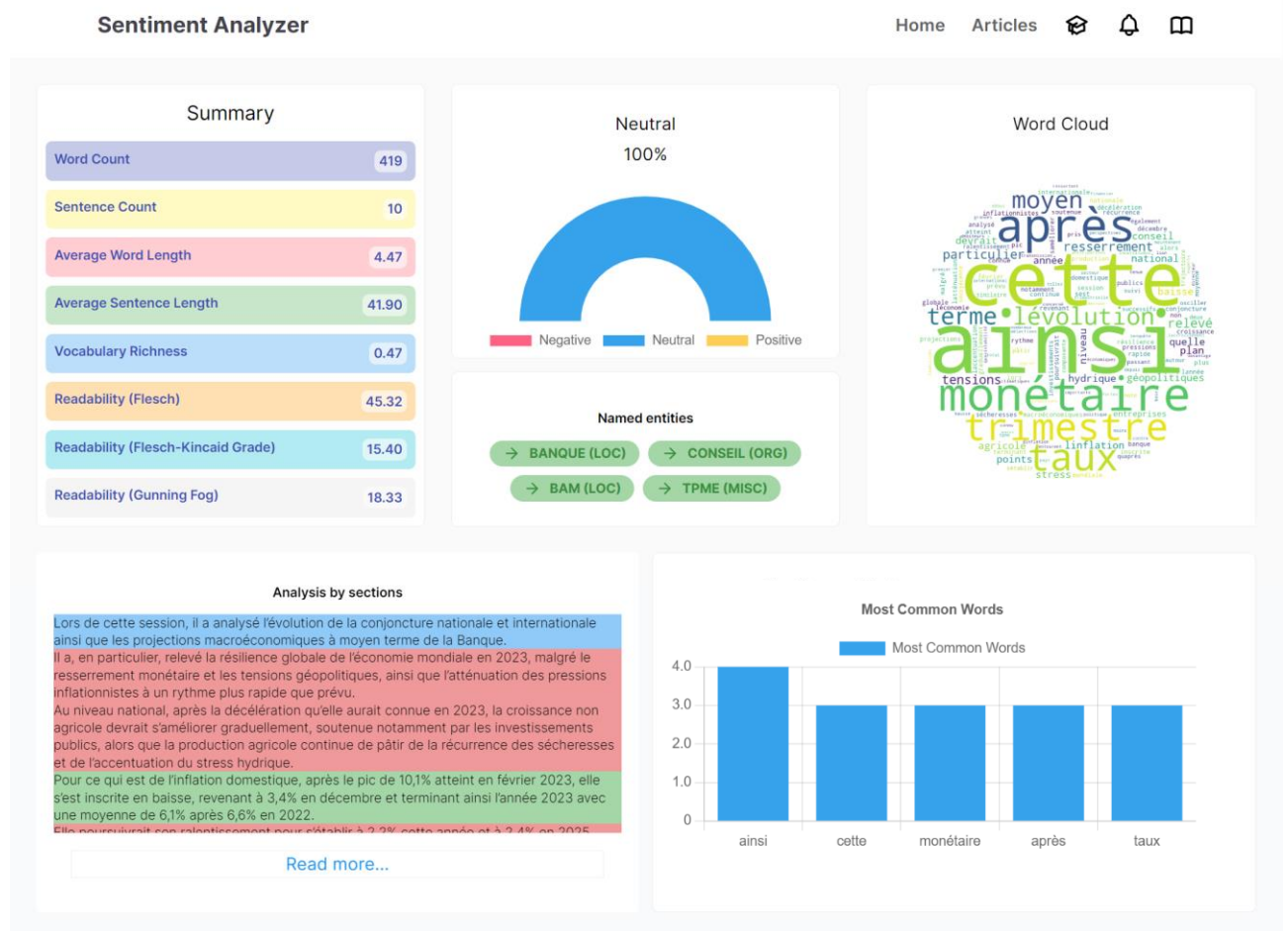


Figure 17: composants de la table de bord

3. Conclusion

À travers un tableau de bord interactif, nous avons pu illustrer de manière visuelle les résultats de nos modèles, y compris les analyses de sentiment, les statistiques des articles, et les visualisations comme les nuages de mots et les diagrammes. Cette interface constitue un outil précieux pour les décideurs et les chercheurs, offrant une vue d'ensemble claire et détaillée sur les perceptions du public concernant la politique monétaire marocaine

Conclusion générale

Ce rapport documente le projet intitulé " l'analyse de sentiments des textes relevant des décisions de la politique monétaire de BANK AL-MAGHRIB ", mené sur une durée de cinq mois. Nous avons initialement commencé par une compréhension approfondie du contexte de la politique monétaire marocaine et du rôle de Bank Al-Maghrib. La banque centrale joue un rôle crucial dans la stabilisation économique et la régulation des taux d'intérêt au Maroc.

Notre projet s'est articulé autour de plusieurs étapes clés, chacune avec ses propres défis et apprentissages. Nous avons débuté par la collecte et le prétraitement des données, un aspect fondamental pour travailler avec des textes non structurés. Les projets de traitement automatique du langage naturel (TALN) sont notoirement complexes en raison de la nature imprévisible et variée des données textuelles. Le nettoyage, la normalisation et la transformation des textes en formats exploitables sont des tâches laborieuses mais essentielles.

Nous avons ensuite construit un corpus de données représentatif et équilibré, puis entraîné plusieurs modèles d'apprentissage automatique, dont le modèle Naive Bayes et le Support Vector Machine (SVM), pour l'analyse des sentiments. Bien que nos résultats soient prometteurs, il est apparu clairement que l'amélioration des performances des modèles nécessite davantage de données et un raffinement continu du corpus. L'expansion du corpus de données et l'intégration de sources diversifiées enrichiraient l'analyse et la rendraient plus représentative des opinions publiques.

La création d'une interface utilisateur et d'un tableau de bord interactif a été une étape cruciale pour rendre les résultats de notre analyse accessibles et utiles aux décideurs et aux chercheurs. Cette interface permet de visualiser les résultats de manière claire et intuitive, facilitant ainsi l'interprétation des sentiments et des statistiques extraites des articles de presse.

Tout au long du projet, nous avons rencontré divers défis, notamment la gestion de grandes quantités de données non structurées et la nécessité d'affiner continuellement nos modèles pour obtenir des prédictions précises. Ces défis soulignent la complexité inhérente aux projets de TALN et l'importance d'une approche méthodique et itérative.

À travers ce projet, j'ai eu l'opportunité de travailler sur un projet concret du monde réel, ce qui m'a permis d'appliquer mes connaissances théoriques à des situations pratiques. J'ai acquis une compréhension approfondie des valeurs et des concepts de la politique monétaire qui influencent l'économie marocaine. Travailler sur un pipeline de traitement des données et de modélisation m'a également permis de développer des compétences techniques avancées et une approche structurée de la résolution de problèmes.

En conclusion, ce projet nous a permis de développer une compréhension approfondie des outils et techniques modernes pour l'analyse des sentiments et a ouvert des perspectives intéressantes pour leur application dans le domaine de la politique monétaire. Les recommandations pour l'avenir incluent l'expansion du corpus de données, l'exploration de techniques avancées en TALN et en modélisation, et l'optimisation de l'interface utilisateur. En intégrant ces améliorations, notre système pourra offrir une contribution encore plus significative à l'analyse économique et aux décisions stratégiques au Maroc.

Bibliographie

- [1] B. Al-Maghrib, «1,» [En ligne]. Available: <https://www.bkam.ma/pedagogique/Principales-missions-de-bank-al-maghrib/Quel-est-le-role-de-bank-al-maghrib>. [Accès le 13 February 2024].
- [2] H.-C. a. P. d. Maroc, «2,» Vendredi 19 Août 2011. [En ligne]. Available: https://www.hcp.ma/Systeme-financier-marocain_a841.html. [Accès le 15 Feb 2024].
- [3] «6,» [En ligne]. Available: <https://drive.google.com/file/d/1NV2giJ5uGGhgLO76WM4VQ9Cw7pDC7cil/view?usp=sharing>.
- [4] B. Al-Maghrib, «3,» [En ligne]. Available: <https://drive.google.com/file/d/1JQnyGIIU0NGuGyZeo1iSI5B54sYDtHoj/view?usp=sharing>. [Accès le 09 Feb 2024].
- [5] F. m. Étude économique n° 4 - Natixis, «4,» 5 novembre 2009. [En ligne]. Available: https://www.senat.fr/eco/ec-04/ec-04_mono.html. [Accès le 15 Feb 2024].
- [6] B. Al-Maghrib, «5,» 2023. [En ligne]. Available: <https://drive.google.com/file/d/1gvQwNTdpPEuiy3gIETJ1YAzDylhUXHN8/view?usp=sharing>. [Accès le 20 Feb 2024].
- [7] analyticsvidhya, «Guide to Sentiment Analysis using Natural Language Processing,» analyticsvidhya, [En ligne]. Available: <https://www.analyticsvidhya.com/blog/2021/06/nlp-sentiment-analysis/>. [Accès le Avril 2024].
- [8] analyticsvidhya, «Feature Engineering in NLP,» analyticsvidhya, [En ligne]. Available: <https://www.analyticsvidhya.com/blog/2021/04/a-guide-to-feature-engineering-in-nlp/>. [Accès le 15 Mai 2024].
- [9] analyticsvidhya, «Naive Bayes Algorithm: Theory & Implementation,» analyticsvidhya, [En ligne]. Available: <https://www.analyticsvidhya.com/blog/2021/01/a-guide-to-the-naive-bayes-algorithm/>. [Accès le 12 Avril 2024].
- [10] analyticsvidhya, «Using Support Vector Machines (SVM) for Data Science,» [En ligne]. Available: <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>. [Accès le 3 Mai 2024].
- [11] «Next.js Documentation,» Next.js, [En ligne]. Available: <https://nextjs.org/docs>. [Accès le 2 Juin 2024].
- [12] «Flask Documentation,» [En ligne]. Available: <https://flask.palletsprojects.com/en/3.0.x/>. [Accès le 2 Juin 2024].

Annexes

Annexe I

```
!pip install unicode

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import nltk
import seaborn as sns
import re

from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer
from sklearn.metrics import accuracy_score, balanced_accuracy_score, precision_score, confusion_matrix, ConfusionMatrixDisplay
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB

from nltk.stem import WordNetLemmatizer
from nltk.stem import PorterStemmer
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords

from unicode import unicode
```

Annexe II

```
# Download necessary NLTK resources
nltk.download("punkt")
nltk.download('wordnet')
nltk.download('stopwords')

# Loading sentiment analysis dataset
df = pd.read_excel('input.xlsx', usecols=['Article', 'tons'])
df.columns = ['tons', 'Article']

# Defining additional stopwords
with open('stop_compl.txt') as f:
    stops = [line.strip() for line in f.readlines()]
```

Annexe III

```
# Mapping sentiment labels to numerical values (e.g., 'Positif' -> 2, 'Négatif' -> 0, 'Neutre' -> 1)
label_map = {'Négatif': 0, 'Neutre': 1, 'Positif': 2}
df['target'] = df['tons'].map(label_map)

# Check for duplicate data
df['duplicates'] = np.where(df['Article'].duplicated(keep=False), 1, 0)
# print('Top 5 common duplicate texts are:')
# print(df[df.duplicates == 1].pivot_table(index='Article', values='target', aggfunc='count').sort_values(by='target', ascending=

# Check for missing data
print("Missing data in each column:\n" + str(df.isnull().sum()))
```

Annexe IV

```
# Exploring distribution of dataset

# count plot on single categorical variable
ax = sns.countplot(x = df['tons'], palette = 'rocket')

# add data labels
ax.bar_label(ax.containers[0])

# add plot title
plt.title("Observations by Classification Type")

# show plot
plt.show()
```

Annexe V

```
# Dropping duplicate texts based on results of EDA
df_no_dup = df.drop_duplicates(subset=['Article'], keep='first')
```

Annexe VI

```
# Exploring distribution of dataset after removing duplicates
sns.countplot(x=df_no_dup['tons'], palette='rocket')
plt.title("Observations by Classification Type (No Duplicates)")
plt.show()
```

Annexe VII

```
# Preprocessing function
def text_clean(text, method, rm_stop):
    text = re.sub(r'<.*?>', '', text) # Remove HTML tags
    text = re.sub(r'https?:\./.*[\r\n]*', '', text) # Remove hyperlinks
    text = re.sub(r'\n', "", text) # Remove line breaks
    text = text.lower() # Convert to lowercase

    # Removing stop words
    if rm_stop:
        stop_words = set(stopwords.words('french')).union(stops)
        filtered_tokens = [word for word in word_tokenize(text) if word.lower() not in stop_words]
        text = " ".join(filtered_tokens)

    text = re.sub(r"\d+", "", text) # Remove digits and currencies
    text = re.sub(r'[\$d+\d+\$]', "", text)
    text = re.sub(r'\d+[\.\./-]\d+[\.\./-]\d+', '', text) # Remove dates
    text = re.sub(r"'\b\w'(\w+)", r'\1', text) # remove apostrophes
    text = unicode(text)
    # text = re.sub(r'^\x00-\x7f', r' ', text) # Remove non-ascii
    text = re.sub(r'^\w\s', '', text) # Remove punctuation

    # Lemmatization or stemming
    if method == 'L':
        lemmer = WordNetLemmatizer()
        lemm_tokens = [lemmer.lemmatize(word) for word in word_tokenize(text)]

        return " ".join(lemm_tokens)
    elif method == 'S':
        porter = PorterStemmer()
        stem_tokens = [porter.stem(word) for word in word_tokenize(text)]
        return " ".join(stem_tokens)

    return text
```

Annexe VIII

```
# Original French sentence with accents
sentence_with_accents = "Je déteste cet à été être pluvieux et sombre, c'est absolument épouvantable."

# Convert to unaccented string
print("Original sentence: ", sentence_with_accents)
print("Sentence with lemmatization: ", text_clean(sentence_with_accents, 'L', True))

print("Sentence with stemming: ", text_clean(sentence_with_accents, 'S', True))
```

```
Original sentence: Je déteste cet à été être pluvieux et sombre, c'est absolument épouvantable.
Sentence with lemmatization: deteste cet etre pluvieux sombre est absolument epouvantable
Sentence with stemming: detest cet etr pluvieux sombr est absolu epouvant
```

Annexe IX

```
# Splitting data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(preprocessed_text, df_no_dup['target'], test_size=0.3, random_state=0)

print("X_train:", len(X_train))
print("X_test:", len(X_test))
print("y_train:", len(y_train))
print("y_test:", len(y_test))

print("Unique values in y_train:", pd.unique(y_train))
print("Unique values in y_test:", pd.unique(y_test))

X_train: 42
X_test: 18
y_train: 42
y_test: 18
Unique values in y_train: [0 1 2]
Unique values in y_test: [1 2 0]
```

Annexe X

```
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Function to generate word cloud for a given text
def generate_wordcloud(text):
    # Generating word cloud
    wordcloud = WordCloud(width=800, height=400, background_color='white').generate(text)

    # Displaying the word cloud
    plt.figure(figsize=(10, 5))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis('off')
    plt.show()

# Iterating over each row in the DataFrame
for index, row in df_no_dup.iterrows():
    print("Word cloud for row", index)
    generate_wordcloud(preprocessed_text[index])

from sklearn.metrics import accuracy_score, balanced_accuracy_score, precision_score, recall_score, f1_score
from nltk.sentiment import SentimentIntensityAnalyzer
import nltk
nltk.download('vader_lexicon')
def transform_and_train_model(preprocessed_text, y_train, X_test, y_test, vectorizer_type='tfidf', ngram_range=(1, 1)):
    if vectorizer_type == 'count':
        vectorizer = CountVecorizer(ngram_range=ngram_range)
    elif vectorizer_type == 'tfidf':
        vectorizer = TfidfVectorizer(ngram_range=ngram_range)

    X_train_vect = vectorizer.fit_transform(preprocessed_text)
    X_test_vect = vectorizer.transform(X_test)

    # Training Multinomial Naive Bayes model
    model = MultinomialNB(alpha=0.1)
    model.fit(X_train_vect, y_train)

    # model evaluation
    predictions = model.predict(X_test_vect)
    accuracy = accuracy_score(y_test, predictions)
    balanced_accuracy = balanced_accuracy_score(y_test, predictions)
    precision = precision_score(y_test, predictions, average='weighted')
    recall = recall_score(y_test, predictions, average='weighted')
    f1 = f1_score(y_test, predictions, average='weighted')

    print("Accuracy:", round(100 * accuracy, 2), '%')
    print("Balanced accuracy:", round(100 * balanced_accuracy, 2), '%')
    print("Precision:", round(100 * precision, 2), '%')
    print("Recall:", round(100 * recall, 2), '%')
    print("F1-score:", round(100 * f1, 2), '%')

    # Calculating sentiment scores for each text in X_test
    sid = SentimentIntensityAnalyzer()
    for i, (text, actual_sentiment) in enumerate(zip(X_test, y_test)):
        sentiment_scores = sid.polarity_scores(text)
        # print(text)
        print(f"Sentiment of text {i + 1}: {sentiment_scores} , actual: {actual_sentiment}")

    return model, predictions
```

Annexe XI

```
# Applying CountVectorizer with Lemmatization and stopwords removal
print("CountVectorizer with Lemmatization and stopwords removal:")
transform_and_train_model(X_train, y_train, X_test, y_test, vectorizer_type='count')

# Applying TfidfVectorizer with Lemmatization and stopwords removal
print("\nTfidfVectorizer with Lemmatization and stopwords removal:")
transform_and_train_model(X_train, y_train, X_test, y_test, vectorizer_type='tfidf')

CountVectorizer with Lemmatization and stopwords removal:
Accuracy: 72.22 %
Balanced accuracy: 69.05 %
Precision: 75.69 %
Recall: 72.22 %
F1-score: 71.85 %

TfidfVectorizer with Lemmatization and stopwords removal:
Accuracy: 61.11 %
Balanced accuracy: 59.52 %
Precision: 60.19 %
Recall: 61.11 %
F1-score: 59.81 %
```

Annexe XII

```
# the confusion matrix
model, predictions = transform_and_train_model(X_train, y_train, X_test, y_test, vectorizer_type='count')
cm = confusion_matrix(y_test, predictions)
disp = ConfusionMatrixDisplay(confusion_matrix=cm, display_labels=model.classes_)
disp.plot()
plt.show()
```

Annexe XIII

```
def transform_and_train_svm(preprocessed_text, y_train, X_test, y_test, vectorizer_type='tfidf', ngram_range=(1, 1)):
    # Vectorization
    if vectorizer_type == 'count':
        vectorizer = CountVectorizer(ngram_range=ngram_range)
    elif vectorizer_type == 'tfidf':
        vectorizer = TfidfVectorizer(ngram_range=ngram_range)

    X_train_vect = vectorizer.fit_transform(preprocessed_text)
    X_test_vect = vectorizer.transform(X_test)

    # Model Training
    model = SVC(kernel='linear', decision_function_shape='ovr') # One-vs-Rest strategy
    model.fit(X_train_vect, y_train)

    # Model Evaluation
    predictions = model.predict(X_test_vect)
    accuracy = accuracy_score(y_test, predictions)
    balanced_accuracy = balanced_accuracy_score(y_test, predictions)
    precision = precision_score(y_test, predictions, average='weighted')
    recall = recall_score(y_test, predictions, average='weighted')
    f1 = f1_score(y_test, predictions, average='weighted')

    # Printing Metrics
    print("SVM Accuracy:", round(100 * accuracy, 2), '%')
    print("SVM Balanced accuracy:", round(100 * balanced_accuracy, 2), '%')
    print("SVM Precision:", round(100 * precision, 2), '%')
    print("SVM Recall:", round(100 * recall, 2), '%')
    print("SVM F1-score:", round(100 * f1, 2), '%')

    # Sentiment Analysis
    sid = SentimentIntensityAnalyzer()
    for i, (text, actual_sentiment) in enumerate(zip(X_test, y_test)):
        sentiment_scores = sid.polarity_scores(text)
        print(f"Sentiment of text {i + 1}: {sentiment_scores}, actual: {actual_sentiment}")

    return model, predictions
```

Annexe XIV

Les données bruts – avant prétraitement

A	B	C	D	E	
ALOUA24.MA - epr	لماكرو اقتصادية	2023-07-08 12:17:18	negatif	Monétaire	1 page
FINANCES NEWS HEBDO - epr	sur le crédit	2023-07-21 00:00:00	positif	Réglementation	1 page
AL ITTIHAD AL ICHTIRAKI - v_online	فضية البرلمان .. الحكومة تقر بالنصار مقارنة لحلمي حول التضخم	2023-07-19 17:30:38	negatif	Politique monétaire	1 page
L'OPINION - epr	financière	2023-07-19 00:00:00	positif	Monétaire	2 pages
LA VIE ECO - v_online	Chambre	2023-07-18 13:46:23	positif	Monétaire	1 page
L'ECONOMISTE - v_online	Les résidents bientôt exclus de Booking.com?	2023-07-17 18:06:15	negatif	Crédit	1 page
2M - epr	Fête du Trône: Bank Al-Maghrib émet une pièce commémorative	2023-07-30 20:30:00	positif	Production Fiduciaire	Chroniq
ALMAGHRIBIALYAOU.COM - epr	تزييف عملات: رصد أزيد من 7000 ورقة نقدية مزيفة خلال سنة 2022 (بنك المغرب)	2023-07-30 20:28:28	negatif	Production Fiduciaire	Quart d
ASSABK.COM - epr	حصصة الأسد لـ"الزرقة".. رصد أزيد من 7000 ورقة نقدية مزيفة في 2022	2023-07-30 20:12:42	negatif	Production Fiduciaire	Quart d
JOURNALASWAT.COM - epr	والعالية بالمغرب	2023-07-30 19:30:41	positif	Rapport Annuel	1 page
LE360 - EPR	250 dirhams	2023-07-30 19:05:14	positif	Production Fiduciaire	Demi-pi
IMAROCPRESS.COM - epr	تقرير مرفوع للملك: المغاربة لا يزالون يشعرون حتى الآن بندني جودة التعليم	2023-07-30 18:45:51	negatif	Rapport Annuel	Demi-pi
ALAYAM24.COM - epr	الجدري: تقرير الجواهري تضمن أمورا لا بد من الوقوف عليها وتطويعها	2023-07-30 16:51:57	positif	Rapport Annuel	Demi-pi
L'ECONOMISTE - v_online	Bank Al-Maghrib: Les messages forts de Jouahri	2023-07-30 16:06:36	positif	Rapport Annuel	1 page
MAGHRESS.COM - epr	الجواهري: هوامش تدخل السلطات لتخفيف الغلاء باتت محدودة وانتظارات المغاربة كبيرة	2023-07-30 14:16:20	negatif	Rapport Annuel	1 page
TANJA 7 - v_online	محذرا من سنوات صعبة قادمة	2023-07-29 19:59:34	negatif	Rapport Annuel	Demi-pi
IKTISADKOM.MA - epr	مخاطر البنوك.. القروض معلقة الأداء تفرغ جريس الإئذان	2023-07-29 15:55:53	negatif	Monétaires	Demi-pi
SAHELPRESS.COM - epr	اختلالات خطيرة في استخدام الشيك من قبل زبناء البنوك ؟	2023-07-28 23:26:57	negatif	Moyens De Paiement	Quart d
LE7TV.MA - epr	بنك المغرب يحيي مناسبة عيد العرش ويصدر قطعة نقدية تذكارية من فئة 250 درهم	2023-07-28 15:51:49	positif	Production Fiduciaire	Quart d
MOGADORINFOS.COM - epr	جلالة الملك يعزي أفراد أسرة المرحوم محمد السقاوط	2023-07-28 14:56:35	positif	Conjoncture	Quart d
FINANCES NEWS HEBDO - v_online	normes bâloises	2023-07-28 12:10:42	positif	Monétaire	1 page
FINANCES NEWS HEBDO - epr	populiste	2023-07-28 00:00:00	negatif	Politique monétaire	Chroniq

Annexe XV

Le corpus après prétraitement

A	B
Article	tons
banqu mondial fait preuv resili remarqu face diver choc suit ralentiss marqu du dive	Positif
conseil concurr met gard contr activ spiral inflationnist marocanass codemal conseil	Négatif
monnai national fort resili face choc economiquessur base evolut plusieurs facteur te	Positif
taux interetinfl politiqu monetair forcerment servic croissancela relat entr taux intere	Négatif
reduct inflat m met avant succ politiqu monetair marocainemarrakech directeur de	Positif
baiss inflat montr succ politiqu monetair gouvernemental marocabdellatif almaghrib	Positif
centr etud commercial control inflat rassur investisseurseconomi financel centr etud	Positif
politiqu monetair necessit ciblag inflat selon fmipolitiqu monetair necessit ciblag infl	Neutre
paradox quelqu chose tourn rond royaum anne apr anne pay continu detruir emplc	Négatif
almaghrib met circul partir nouveau billet banqu dirham nouvel seri piec monnai ca	Positif
inflat secheress crise ukrainienn brouillent cart rapport vit situat economiqu difficil c	Négatif
almaghrib lanc virement bancaire instantan gratuit pendant moisl virement bancaire ir	Négatif
connaiss bien savez vieil oncl invit peu retic auxrassembl familiaux nich coin salon m	Négatif
quand institut font travail	Positif
autorit interdis march contr vie chere milit accroch manifest autorit interdit march n	Négatif
previs maintien taux directeur banqu central context ralentiss inflat marocla plupart	Neutre
economist kech decis banqu central augment taux interet cree crise marocpubli par	Négatif
almaghrib surprend march mainten taux directeur inchangel centr etud recherch co	Neutre