

RÉPUBLIQUE DE CÔTE D'IVOIRE



INTERNATIONAL DATA SCIENCE INSTITUTE



PROJET DE TEXT MINING

Analyse Comparative des Discours Sanitaires

RÉALISÉ PAR :

DAGO Kack'reaux Louis Charles

NIADA Ulrich Boris

MOUHI Christ Emmanuel

PÉRIODE :

Avril - Juin 2025

ENSEIGNANT :

DR TANOH LAMBERT

Directeur de l'IDSI

DSI de l'INPHB

MASTER 1 DATA SCIENCE, BIG DATA & INTÉLLIGENCE ARTIFICIELLE

Année académique 2024-2025

Table des matières

0.1	Résumé exécutif	3
0.2	Introduction	3
0.3	Corpus et Description des Données	3
0.3.1	Corpus OMS Afrique	3
0.3.2	Corpus Forbes Afrique	3
0.4	Méthodologie	4
0.4.1	Nettoyage et Prétraitement	4
0.4.2	Modélisation du Texte	4
0.4.3	Extraction des Thèmes OMS (LDA)	4
0.4.4	Analyse du Sentiment et Framing	4
0.4.5	Extraction d’Entités (NER – spaCy)	4
0.5	Résultats	4
0.5.1	Thèmes OMS Identifiés	4
0.5.2	Couverture Forbes des Thèmes OMS	5
0.5.3	Analyse UMAP – Proximité Sémantique	5
0.5.4	Analyse du Sentiment	5
0.5.5	Analyse des Entités et Wordclouds	5
0.6	Conclusion	5

0.1 Résumé exécutif

Ce rapport analyse dans quelle mesure les priorités sanitaires communiquées par l’Organisation Mondiale de la Santé (OMS) – Région Afrique sont reprises, reformulées ou ignorées par Forbes Afrique, média économique majeur du continent.

L’analyse repose sur un pipeline complet de *text mining* incluant :

- extraction de thématiques (LDA);
- similarité lexicale et sémantique (TF-IDF, SBERT);
- analyse du sentiment;
- analyse du *framing* éditorial;
- extraction d’entités nommées (NER);
- visualisation UMAP;
- création d’un dashboard interactif.

Les résultats montrent un écart important entre les enjeux de santé publique (OMS) et la vision économique centrée sur l’innovation (Forbes Afrique).

0.2 Introduction

L’Afrique fait face à de nombreux défis sanitaires : épidémies récurrentes, mortalité infantile élevée, faible couverture vaccinale, infrastructures insuffisantes.

L’OMS documente ces problématiques dans ses communications officielles. Forbes Afrique, en revanche, focalise sa couverture sur l’économie, l’innovation et l’entrepreneuriat.

Objectif général : mesurer le degré de convergence entre ces deux visions à travers une analyse textuelle approfondie.

Questions de recherche :

- Quels sont les thèmes prioritaires de l’OMS ?
- Dans quelle mesure Forbes couvre-t-il ces thèmes ?
- Quelle est la proximité sémantique OMS–Forbes ?
- Comment les deux sources diffèrent-elles en termes de sentiment et d’angle éditorial (*framing*) ?

0.3 Corpus et Description des Données

0.3.1 Corpus OMS Afrique

- ~ 60 articles
- Langue : principalement français
- Types : communiqués, bulletins épidémiologiques, rapports terrain
- Thèmes récurrents : épidémies, vaccination, mortalité maternelle, urgences sanitaires

0.3.2 Corpus Forbes Afrique

- ~ 50 articles
- Langue : français

- Types : interviews, reportages économiques, analyses de marché
- Thèmes : innovation, biotech, start-ups santé, financements

0.4 Méthodologie

0.4.1 Nettoyage et Prétraitement

- normalisation : minuscules, suppression URLs, accents ;
- suppression des mots outils FR/EN ;
- tokenisation ;
- lemmatisation/stemming.

0.4.2 Modélisation du Texte

- **TF-IDF** : comparaison lexicale ;
- **SBERT** (all-mpnet-base-v2) : représentation sémantique ;
- **UMAP** : réduction dimensionnelle.

0.4.3 Extraction des Thèmes OMS (LDA)

- 8 thèmes extraits automatiquement ;
- servent de référence pour mesurer la couverture Forbes.

0.4.4 Analyse du Sentiment et Framing

- modèle Transformer multilingue ;
- labels : POS / NEG / NEU ;
- *framing* : économique, sanitaire, mixte, neutre.

0.4.5 Extraction d'Entités (NER – spaCy)

Labels utilisées : PER, ORG, LOC, GPE, MISC.

0.5 Résultats

0.5.1 Thèmes OMS Identifiés

- Épidémies (Ebola, choléra)
- Vaccination
- Santé maternelle
- Accès aux soins
- Surveillance épidémiologique
- Tuberculose
- Urgences humanitaires

— Renforcement des systèmes de santé

Tableau 1 : Résumé des 8 thèmes OMS extraits. (À compléter avec votre tableau)

0.5.2 Couverture Forbes des Thèmes OMS

À insérer : Barplot couverture Forbes → thèmes OMS

Interprétation (pré-rédigée) : Forbes couvre principalement les thèmes liés à l'innovation, aux investissements et au marché de la santé. Les thématiques critiques (choléra, crises humanitaires, mortalité maternelle) sont très peu relayées.

0.5.3 Analyse UMAP – Proximité Sémantique

Interprétation (pré-rédigée) : L'UMAP montre deux clusters distincts :

- OMS : discours sanitaire, urgence, santé publique ;
- Forbes : innovation, économie, financement.

Quelques points hybrides apparaissent sur les sujets de vaccination et de systèmes de santé.

0.5.4 Analyse du Sentiment

Interprétation : OMS adopte un ton neutre ou négatif (crises sanitaires), tandis que Forbes est majoritairement positif (innovations, success stories).

0.5.5 Analyse des Entités et Wordclouds

Interprétation : OMS : crise, maladie, épidémie, vaccination. Forbes : innovation, technologie, entreprise, marché.

0.6 Conclusion

Cette étude met en évidence un écart significatif entre les priorités sanitaires de l'OMS et les priorités éditoriales de Forbes Afrique.

Résumé des divergences majeures :

- Santé publique vs Santé économique ;
- urgence sanitaire vs innovation ;
- institutions publiques vs acteurs privés.

Perspectives (à compléter) :

- Analyse temporelle ;
- Analyse par pays ;
- Inclusion d'autres médias africains.