

## Résumé de l'article Mining news articles dealing with Food Security

L'article évoque un projet visant à améliorer la sécurité alimentaire en Afrique de l'Ouest, en mettant l'accent sur le Burkina Faso, une région touchée par une crise humanitaire depuis le début du siècle. L'approche proposée consiste à combiner l'analyse de données textuelles, telle que l'utilisation de l'embarquement de mots, l'analyse de sentiment et le calcul de la discrimination, avec les données numériques traditionnelles pour obtenir un modèle explicatif. Ce modèle est évalué sur des données réelles à grande échelle et offre des informations qualitatives complémentaires sur la sécurité alimentaire et ses caractéristiques spatiales et temporelles. Un accent important est sur la persistance de la faim dans de nombreuses régions du monde et de la situation critique au Burkina Faso, où la prévalence de la sous-alimentation est élevée. On explore ensuite le concept de « triple fardeau de la malnutrition » qui regroupe la suralimentation, la sous-alimentation et les carences en micronutriments.

Dans l'optique de résoudre ce problème, il s'est avéré nécessaire de créer des systèmes d'alerte et de surveillance de la sécurité alimentaire. Ces systèmes visent à prévenir des crises alimentaires futures afin d'optimiser la production et la distribution alimentaire. On se penche ensuite sur l'utilisation de l'exploration de texte pour extraire et analyser des informations qualitatives à partir de journaux burkinabés, afin de fournir des indicateurs explicatifs complémentaires aux scores de sécurité alimentaire générés automatiquement.

On examine les défis liés à l'utilisation d'approches d'exploration de texte pour analyser la sécurité alimentaire, en mettant l'accent sur le domaine de l'agriculture. Plusieurs études ont abordé des aspects tels que l'analyse de sentiment et l'extraction d'entités nommées dans des textes liés à l'agriculture. On présente une approche multidisciplinaire qui combine l'extraction de texte, l'analyse de sentiment et l'analyse spatiotemporelle pour étudier la sécurité alimentaire, un angle peu exploré jusqu'à présent. L'approche vise à fournir un cadre explicatif complémentaire aux modèles prédictifs classiques, notamment pour le cas du Burkina Faso, tout en étant générique et applicable à d'autres régions. La section 2 décrit l'approche proposée, tandis que la section 3 présente les résultats de l'analyse de données textuelles en français.

Des méthodes d'exploration de texte ont été utilisées pour extraire des informations sur la sécurité alimentaire à partir de journaux. Ils utilisent la méthode tf-idf pour extraire des mots-clés caractéristiques et un modèle Bi-LSTM-CNN-CRF pour les entités nommées. Ces mots-clés sont pondérés en fonction de leur similarité sémantique avec le titre de l'article. Les articles sont ensuite classés en utilisant ces mots-clés et entités. Ils intègrent également Word2vec pour transformer les termes en vecteurs. Cette approche vise à détecter automatiquement les crises alimentaires à partir de données textuelles de journaux.

Deux concepts clés sont abordés :

1. **Polarité des termes** : Il s'agit de déterminer si un terme est positif, négatif ou neutre. Dans le contexte de la sécurité alimentaire, la polarité moyenne des textes peut indiquer leur caractère inquiétant. Pour évaluer la négativité des termes dans des textes en français, le modèle VADER est utilisé. Ce modèle repose sur un lexique de termes classés en positif ou négatif, ainsi que sur des règles contextuelles.
2. **tf-idf** : Cette méthode évalue l'importance d'un terme dans un texte en examinant sa fréquence dans le texte et sa distribution dans l'ensemble des textes étudiés. Elle permet de déterminer si un terme est caractéristique d'un texte.

Le processus d'analyse spatiotemporelle de la sécurité alimentaire comprend trois étapes principales :

1. Sélection des articles pertinents en utilisant la similarité sémantique par rapport à un lexique général (GLEX) pour identifier les articles liés à la sécurité alimentaire.
2. Création de proxies textuels pour évaluer le degré de connexion des articles à la sécurité alimentaire, leur taux de négativité, et l'utilisation de vocabulaire spécifique lié à la sécurité alimentaire et aux crises.
3. Agrégation des proxies à différents niveaux (global, régional, annuel) pour analyser la sécurité alimentaire sur une période donnée. Cette approche permet de détecter les tendances et les crises alimentaires.

Des expériences ont été menées en utilisant un corpus de journaux burkinabés pour analyser la sécurité alimentaire de 2009 à 2018. Il se concentre sur une analyse régionale en examinant la corrélation entre les proxies de sécurité alimentaire et la situation alimentaire régionale, en mettant en évidence des différences significatives. Ensuite, il effectue une analyse annuelle pour voir comment les proxies de sécurité alimentaire ont évolué au fil du temps. Les résultats montrent des tendances, telles que la diminution des articles négatifs au fil des ans, et des pics liés à des événements spécifiques. Ces analyses fournissent des informations importantes sur la sécurité alimentaire au Burkina Faso.

Dans cette étude, nous avons exploré l'utilisation de méthodes d'analyse de texte pour extraire des informations thématiques sur la sécurité alimentaire à partir d'articles de journaux en se concentrant sur la situation au Burkina Faso. Nous avons proposé trois types de proxies basés sur des méthodes d'analyse de texte adaptées, notamment le modèle d'encastrement lexical Word2vec, le modèle d'analyse de sentiment VADER et la méthode de pondération de l'importance des termes tf-idf. Ces proxies nous ont permis d'obtenir des informations complémentaires et distinctes sur le thème de la sécurité alimentaire. Cette approche peut être utilisée pour compléter les résultats de modèles prédictifs, en particulier ceux basés sur l'apprentissage automatique, qui souvent manquent d'explicabilité. Pour améliorer encore la

recherche thématique, des technologies basées sur BERT et des modèles entraînés pour le français pourraient être intégrés.