Review of the use of text mining in food price prediction and food security models: A Semantic AI approach

Ghada Alaa
AI Principal Lead - Information Sector
Information & Decision Support Center (IDSC) - Egypt

Ghada Refaat El Said
Department of Management Information Systems
Future University in Egypt (FUE), Egypt

Keywords
Food Price Forecasting; Food Security; Text Mining; Natural Language Processing; Sentiment Analysis; LSTM

Abstract

The increasing volatility of food markets requires more responsive and accurate forecasting models to support effective policy decisions and ensure adaptation to the market. While traditional forecasting methods in agriculture rely heavily on structured quantitative data, they often overlook the latent predictive signals embedded in unstructured textual sources such as news, social media posts, government reports, and expert insights. Same, for food security value chain that covers primarily food production, distribution, and consumption, text mining techniques can help identify trends, sentiment polarity, political and market directions.

This paper systematically reviews research in the domain of integrating text mining techniques with quantitative food price prediction and food security models. We conduct a structured literature search across multiple academic databases, applying inclusion criteria that emphasize methodological transparency, empirical validation, and relevance. To extract dominant research themes, we used generative AI tool to summarize the research papers we collected. Informative summaries were generated. Then we employed a topic modeling approach that builds on sentence-BERT embeddings and k-means clustering. This provided grouping of issues found in the research papers while ensuring rich semantic relevance. Then with the use of generative AI tool, titles were generated for each group, these represent thematic research themes. Results highlight the growing use of LSTM and hybrid forecasting models, topic modelling, sentiment analysis and domain-specific transformers in agri-food modelling. With more emphasis on sentiment analysis for food price forecasting, and more emphasis on information extraction, text clustering and trend analysis for food security. Our findings underscore the potential of text-augmented models to enhance the timeliness, contextual awareness, and accuracy of food price forecasts and food security models.

Introduction

Accurate forecasting of agricultural prices and food security indicators are a cornerstone of effective policymaking, market regulation, and humanitarian planning. Traditional quantitative models, while robust in handling structured data such as historical prices and climatic variables, often fall short in capturing the dynamic, real-time signals embedded in unstructured textual data. Text mining techniques, that cover Natural Language Processing (NLP) and recently large language model architectures, provide several benefits when integrated with food price prediction. They incorporate into the forecasting process, qualitative signals from news, articles, social media posts, government reports and expert feedback.

Recent research has demonstrated that textual data can significantly enhance the predictive and adaptive power of time series models. For instance, sentiment analysis of agricultural news and social media posts has been shown to improve commodity price forecasts by capturing market expectations and behavioral cues (Xu & Hsu, 2022). Similarly, topic modeling and semantic analysis have been used to detect early warnings of supply chain disruptions and policy shifts that influence food prices (Wihartiko et al., 2021).

In the agricultural domain, the application of NLP is gaining momentum. Research has explored the use of transformer-based models such as BERT and its variants for tasks ranging from crop disease detection to semantic classification of agricultural documents (Reimers & Gurevych, 2019). Moreover, hybrid models that combine deep learning architectures with text-derived features, such as keyword vectors or sentiment scores have shown measurable improvements in forecasting accuracy for agricultural exports and inflation trends (Xu & Hsu, 2022).

Despite these advances, the integration of text-based insights into quantitative forecasting remains underexplored in agriculture compared to domains like finance or marketing. A recent review by Wihartiko et al. (2021) highlighted that while deep learning models and data mining techniques are increasingly used in agricultural price prediction, the role of unstructured text data is still emerging. Similarly, the Frontiers Research Topic on Semantics and NLP in Agriculture underscores the untapped potential of semantic technologies in addressing agri-business concerns through text analytics.

This paper aims to fill this gap by systematically reviewing the literature on integrating text-based methods in food price prediction and food security models. Through this synthesis, we aim to map the current methodological landscape, identify key NLP techniques and architectures in use for agri-business analysis, and inform future work in this interdisciplinary field.

Literature Review

Food Price Forecasting

The increasing volatility of food prices, driven by global disruptions, climate change, and market speculation, has prompted researchers to explore advanced forecasting techniques. Among these, text-based forecasting methods have gained prominence for their ability to incorporate unstructured data such as news articles, social media posts, and search trends into predictive models. This section reviews and compares recent studies that integrate textual data into food price forecasting, highlighting methodological innovations, data sources, and forecasting performance.

Jayadianti et al. (2023) focused on forecasting staple food prices in Indonesia using Long Short-Term Memory (LSTM) networks trained on three years of historical price data. Their model demonstrated strong generalizability across ten commodities and was proposed as a decision-support tool for regional inflation control. Notably, this study did not incorporate textual data, serving as a baseline for comparison with more complex, text-integrated models. In contrast, Zhang et al. (2025) introduced a hybrid model combining Graph Convolutional Networks (GCN), Bidirectional Gated Recurrent Units (BiGRU), and Tree-structured Parzen Estimators (TPE). This model captured both structural and temporal dependencies in agricultural futures data, achieving high predictive accuracy ($R^2 = 0.969$). Although it did not explicitly use textual data, its architecture reflects the trend toward multi-dimensional modeling.

Several studies explicitly integrated textual data. Chuluunsaikhan et al. (2020) used Latent Dirichlet Allocation (LDA) to extract topics from South Korean news articles related to pork, which were then used to train LSTM models. The study found a strong correlation between news content—especially disease-related events—and pork price fluctuations, demonstrating the predictive value of media sentiment. Wang and Liu (2025) advanced this approach by combining BERTopic and FinBERT to extract topic and sentiment features from Chinese news headlines about soybean futures. These features were fused with market data and modeled using a hybrid HMM-LSTM architecture (the application of Hidden Markov Models/HMM and Long Short-Term Memory (LSTM) neural networks). The model outperformed baselines, particularly during periods of market disruption, highlighting the value of sentiment-aware forecasting.

Social media data also featured prominently. Inoue and Nakajima (2024) analyzed tweets to assess consumer awareness of New Zealand kiwifruit in Japan. Using text regression and time-series analysis, they demonstrated that social media sentiment significantly influenced export performance. Similarly, Wang et al. (2024) used Weibo data and ChineseBERT to construct a sentiment index for corn futures forecasting. Their model incorporated signal decomposition and SHAP analysis, achieving high accuracy over 30- and 60-day horizons.

An et al. (2024) proposed a multi-modal ensemble model that integrated news text, Baidu search trends, and historical prices to forecast soybean futures. A sentiment intensity index reflecting investor concern was embedded into a multi-task autoencoder, yielding superior performance in both short- and medium-

term forecasts. Li et al. (2022) developed a framework using DP-Sent-LDA and Bi-LSTM to extract sentiment and topic features from online news headlines. These were input into SVR and BPNN models, which outperformed traditional benchmarks in forecasting soybean futures.

Commonalities across these studies include the widespread use of LSTM-based architectures, the integration of sentiment and topic modeling techniques, and a focus on commodities like soybean, corn, and pork. Most studies demonstrated that incorporating textual data—especially sentiment features—significantly improves forecasting accuracy. Discrepancies arise in data sources and modeling complexity. While some studies relied solely on historical price data (e.g., Jayadianti et al., 2023), others integrated multi-source textual data (e.g., An et al., 2024). The sophistication of text analysis also varied, from basic LDA (Chuluunsaikhan et al., 2020) to advanced models like FinBERT and BERTopic (Wang & Liu, 2025). Furthermore, while some models focused on short-term forecasting, others addressed medium- and long-term horizons. The following section analyze the methodologies, data sources, and findings across the reviewed studies.

Model Architectures and Hybrid Techniques:

Following is a comparison of the machine learning and deep learning models used across studies, emphasizing hybrid and ensemble approaches.

- Text Processing Techniques: Topic modeling (LDA, BERTopic), sentiment analysis (FinBERT, Bi-LSTM), and text regression.
- Deep Learning Models: LSTM (Jayadianti et al., 2023; Chuluunsaikhan et al., 2020), BiGRU (Zhang et al., 2025), Bi-LSTM (Li et al., 2022).
- Hybrid Models: GCN–BiGRU–TPE (Zhang et al., 2025), HMM-LSTM (Wang & Liu, 2025), and multi-modal ensemble models (An et al., 2024).
- Optimization Techniques: Use of Optuna for hyperparameter tuning and SHAP for interpretability (Wang et al., 2024).

Sentiment and Topic Integration:

It is important to note how sentiment and topic features are extracted and integrated into forecasting models.

- Sentiment Index Construction: Raw vs. denoised sentiment indices (Wang et al., 2024), investor concern-based sentiment (An et al., 2024).
- Topic Modeling: LDA (Chuluunsaikhan et al., 2020), DP-Sent-LDA (Li et al., 2022), BERTopic (Wang & Liu, 2025).
- Impact on Forecasting: Studies consistently show that integrating sentiment and topic features improves short- and medium-term forecasting accuracy.

Data Sources and Market Contexts:

In the following we show geographical and market contexts of the studies and the types of commodities forecasted, and the sources of textual data used.

- Geographical Focus: China (Wang et al., 2024; Li et al., 2022), South Korea (Chuluunsaikhan et al., 2020), Indonesia (Jayadianti et al., 2023), Japan (Inoue & Nakajima, 2024).
- Commodity Types: Pork, corn, soybean, and general food commodities.
- Data Types: Historical price data, news articles, social media content, and search engine trends.
- Textual Sources: News headlines (Li et al., 2022; Wang & Liu, 2025), tweets (Inoue & Nakajima, 2024), and social media platforms like Weibo (Wang et al., 2024).

Challenges and Research Gaps

Limitations and areas for future research are highlighted as follows.

- Challenges:
- Limited generalizability across regions and commodities.
- Difficulty in quantifying non-market factors
- Lack of real-time data integration.

- Research Gaps:
- Underexplored commodities and regions.
- Need for explainable AI in forecasting.
- Integration of multimodal data (e.g., images, satellite data).

Food Security

Recent scholarship on food security reflects a growing convergence around the use of advanced technologies and interdisciplinary approaches to address global food challenges. A significant body of work emphasizes the role of machine learning (ML) and artificial intelligence (AI) in enhancing food system resilience. Jarray et al. (2023) and AbdulKader et al. (2024) provide comprehensive reviews of ML and remote sensing applications in crop yield prediction, land mapping, and sustainable agriculture, aligning with SDG-2. These studies underscore the potential of AI to process heterogeneous data and support decision-making in rural ecosystems. Complementing this, Reddy et al. (2024) integrate healthcare data modeling with food safety analysis, highlighting the intersection of public health and food security. Their work demonstrates how deep learning can detect food adulteration and trace foodborne illness patterns, offering a novel perspective on urban food safety.

Text mining emerges as another critical tool. Xiong et al. (2024) and Vasilescu et al. (2024) explore its application in food quality control, policy analysis, and public consultation. Ma and Zheng (2025) apply sentiment analysis to social media data, revealing how public opinion can influence food safety responses during crises like the "Rat-Headed Duck Neck" incident. Economic dimensions are addressed by Silva et al. (2024) and Pratap et al. (2022), who develop nowcasting models and sentiment-based forecasting tools to monitor food price inflation. These models are vital for timely policy interventions, especially in volatile markets.

In contrast, Leroy et al. (2015) focus on household-level food access, advocating for experience-based and dietary diversity indicators. Their work highlights the need for low-burden, validated tools for large-scale surveys. While all studies converge on the importance of data-driven approaches, they differ in scale, methodology, and focus, ranging from macroeconomic modeling to micro-level access indicators. This diversity reflects the complexity of food security and the need for integrated, context-sensitive solutions.

Technological Integration and Innovation

A major convergence across studies is the application of AI, ML, and remote sensing to enhance food security. Jarray et al. (2023) and AbdulKader et al. (2024) both emphasize the use of ML for crop yield prediction and sustainable agriculture. However, while Jarray et al. focus on a systematic review of ML models across global contexts, AbdulKader et al. propose a specific GIS-based model tailored to rural ecosystems, achieving high predictive accuracy. Reddy et al. (2024) extend this technological lens by integrating healthcare data with food safety monitoring, using deep learning to detect food adulteration and trace foodborne illnesses. This study uniquely bridges public health and food security, unlike the others which focus more on agricultural productivity.

Public Sentiment and Communication

Ma and Zheng (2025), Xiong et al. (2024), and Vasilescu et al. (2024) explore the role of text mining and sentiment analysis. Ma and Zheng analyze Weibo data during a food safety crisis, showing how public sentiment can escalate or mitigate food-related emergencies. Xiong et al. provide a broader survey of text mining applications in food quality and personalization, while Vasilescu et al. use quantitative text analysis to map stakeholder positions in food policy debates. The common thread is the use of unstructured textual data to inform food safety and policy, but they differ in scope: Ma and Zheng focus on crisis response, Xiong et al. on industry applications, and Vasilescu et al. on regulatory influence.

Economic Monitoring and Price Volatility

Silva et al. (2024) and Pratap et al. (2022) address food price inflation using real-time data. Silva et al. develop nowcasting models to detect inflationary trends, while Pratap et al. incorporate media sentiment to improve forecasting accuracy for volatile commodities like tomatoes and onions. Both studies highlight

the importance of early warning systems, but Pratap et al. uniquely integrate climate and media data, offering a more holistic model.

Measurement and Indicators of Food Access

Leroy et al. (2015) take a different approach by focusing on household-level food access indicators. They critique existing metrics and recommend experience-based and dietary diversity indicators for large-scale surveys. Unlike the tech-heavy studies, this work emphasizes low-burden, validated tools for assessing food access, particularly in vulnerable populations.

Methodological and Contextual Differences

- Scale: Some studies operate at a macro level (e.g., Jarray et al., Silva et al.), while others focus on micro or household levels (e.g., Leroy et al.).
- Data Types: There's a divide between studies using structured data (e.g., satellite imagery indicators, price indices) and those leveraging unstructured data (e.g., social media, policy texts).
- Outcomes: While some aim to predict and prevent (e.g., crop insufficiency, price spikes), others aim to understand and respond (e.g., public sentiment, policy influence).

Research Methodology

To systematically identify and analyze relevant literature on text-based food forecasting methods, this review employed a structured search strategy across multiple academic databases. Key sources included ScienceDirect, Elsevier, Wiley, SpringerLink, Sage, ResearchGate, and Google Scholar, chosen for their comprehensive coverage of peer-reviewed journals and conference proceedings. The search process utilized Boolean operators and targeted keyword combinations to ensure relevance and precision, including the following:

- "text mining" AND "food price prediction"
- "natural language processing" AND "agricultural price forecasting"
- "news sentiment" AND "food price forecasting"
- "social media" AND "food price prediction"
- "text-based" AND "commodity prices"

The food analysis domain does not only cover price predictions, but it handles issues related to food security in general. This covers market analysis, early warning systems for food security crises, such as famines or foodborne disease outbreaks, etc. Also, estimations of inflation levels are a common concern in food security. This made us search for text-based methods in food security analysis, beside food price prediction. We used Boolean operators and targeted keyword combinations to ensure relevance and precision, including the following:

- "text mining" AND "food security"
- "natural language processing" AND "food security"
- "sentiment" AND "food security"
- "social media" AND "food security"
- "text-based" AND "inflation"
- "sentiment" AND "inflation"

Inclusion criteria focused on studies that integrated textual data, such as news articles, social media posts, or government reports, into forecasting models and food security analysis. Studies that lacked methodological transparency were excluded. This approach ensured a focused and high-quality selection of literature, enabling a comprehensive synthesis of current methodologies in the domain under study.

To detect the main topics discussed in the research papers, we applied a topic modelling approach based on large language model embeddings; Sentence-BERT (SBERT) (Reimers & Gurevych, 2019). SBERT is a modification of the pretrained BERT network to derive semantically meaningful sentence embeddings that can be compared using cosine-similarity. It reduces the time for finding the most similar pair from 65 hours with BERT to about 5 seconds with SBERT, while maintaining accuracy. In this research we used the 'all-MiniLM-L6-v2' model, a powerful SBERT representation capable of mapping sentences and paragraphs to a 384-dimensional dense vector space that captures the semantic information of the input. It is useful for

tasks like information retrieval, clustering, and sentence similarity detection. Unlike other text embeddings such as word2vec and Glove, SBERT enriches word representations with contextual information.

The research approach adopted in this study includes (1) collecting the research papers with the criteria discussed above to ensure methodical underpinning, (2) generating informative summaries for the research papers using generative AI tool (600 words-length summaries were found suitable to reflect the research content), (3) converting the research summaries into sentence embedding; 'all-MiniLM-L6-v2' was chosen, (4) clustering the sentence embeddings into groups, (5) generating titles to each group using generative AI tool, and finally (6) retrieving from the research papers most relevant paragraphs related to the titles generated; for this task 'all-mpnet-base-v2' model was used.

The all-SBERT models were trained on all available training data (more than 1 billion training pairs). The 'all-mpnet-base-v2' model provides the best quality, as it maps sentences to a 768-dimensional dense vector space, in contrast to 'all-MiniLM-L6-v2' that uses 384-dimensional vector representation. In retrieving relevant paragraphs related to the generated cluster titles we used 'all-mpnet-base-v2' as we wanted to ensure richness of semantic representation and most accurate results, whereas in the clustering part we used a less dimensional representation 'all-MiniLM-L6-v2' to ensure generalizability and high level abstraction that is required for the clustering task.

The clustering algorithm we used was k-means with a maximum of 300 iterations. We calculated inertia values to decide the optimal number of clusters. Compared to other clustering methods, such as density and hierarchical clustering, k-means produces more coherent and well-distributed topics. As a result, we obtained clusters corresponding to different topics discussed in the research papers.

For these clusters we generated titles using generative AI tool, and for each title we extracted the most relevant paragraphs from the research papers through semantic information retrieval. This resulted in a 'roadmap that reflects current research status and concerns in the field of integrating text-based analysis with quantitative food price prediction and food security models'. As the analysis is carried out using well-proven sentence embeddings, it is more accurate than the human being to depict the recurrent research themes. We also used the state-of-the-art in sentence embeddings and generative AI tools that produce summaries and titles. This made the analysis more informative than just ending up with most frequent words or expressions in the research papers. Furthermore, we calculated the TFIDF score for the text included in each cluster, this gave us the weight for each cluster according to the perspective of the papers.

The thematic titles reflect the research concerns handled in the papers, and implicitly cover the buzzwords for this research domain. The buzzwords are the nouns in the titles, and we calculated a normalized repetition score for each buzzword by dividing the number of its occurrence by the number of characters of the papers it appeared in.

Results

After we followed the methodology explained in previous section, the generative AI tool provided the titles for both food price prediction and food security clustered themes (Figure1 and Figure2). Using the semantic information retrieval approach explained before, we extracted the most relevant paragraphs from the papers that match these titles/themes. In the following we provide the major issues found for these thematic titles.

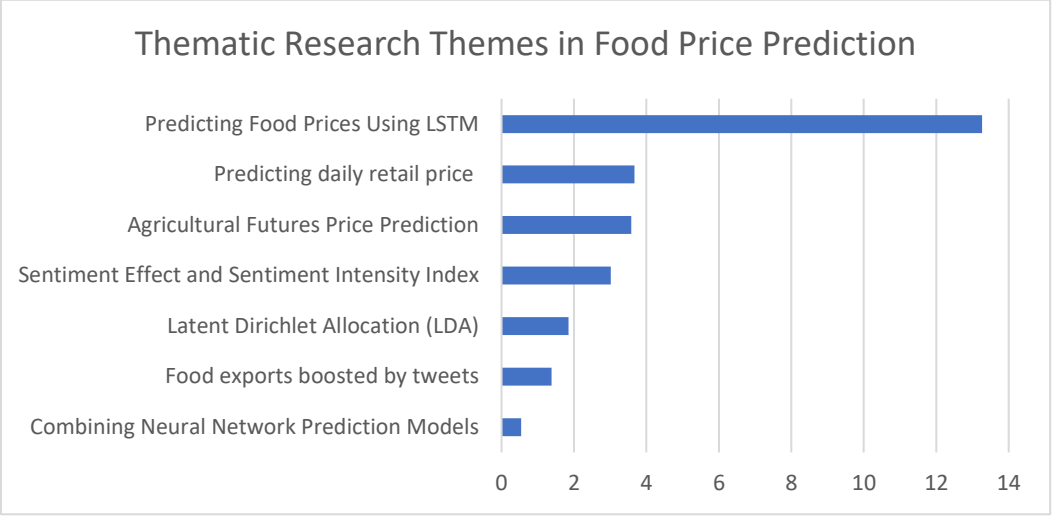

Figure1: Thematic Research Themes for Food Price Prediction

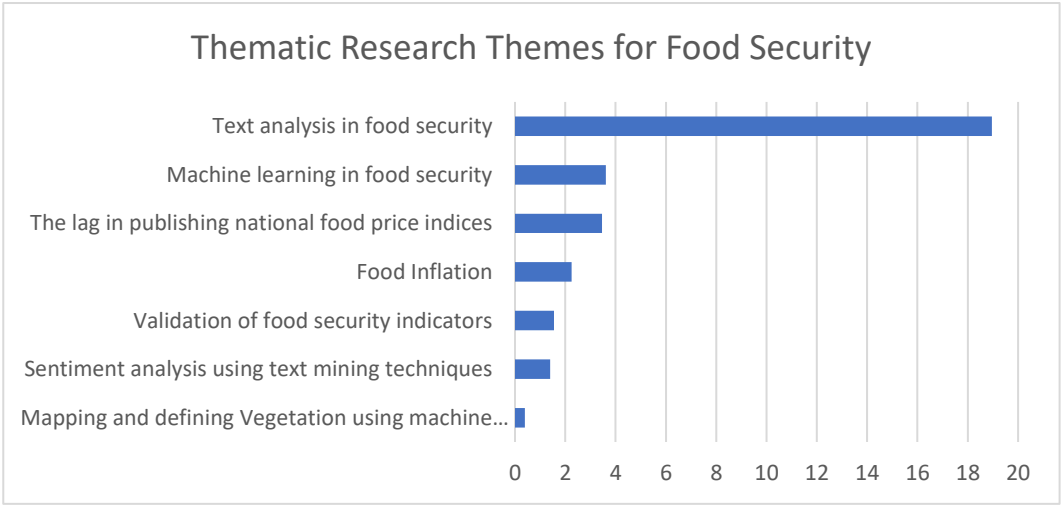

Figure2: Thematic Research Themes for Food Security

Food Price Forecasting Themes

Predicting Food Prices Using LSTM

– LSTM forecast can be used to develop a more accurate and reliable food price prediction model (Jayadianti et al., 2023)

– The outcome of this research is the implementation of an LSTM model architecture for price forecasting and analysis of ten food commodities (Jayadianti et al., 2023)

– The results demonstrated that the proposed LSTM architecture model was generalizable to all commodities and performed well in most cases (Jayadianti et al., 2023).

– The Bi-LSTM model excels in extracting information related to the order of words in a sentence, and thus is used in sentiment analysis (An et al., 2024).

Predicting daily retail price

– There are four types of price data, i.e., the retail price, marketplace price, distribution price, and auction price (Chuluunsaikhan et al., 2020)

– A methodology that predicts the daily retail price of pork in the South Korean domestic market. They utilized news articles and retail price data. Initially a topic modeling technique is carried out to obtain relevant keywords that express price fluctuations. Based on these keywords, a prediction models is constructed using statistical, machine learning, and deep learning methods (Chuluunsaikhan et al., 2020).

Agricultural Futures Price Prediction

– The accurate forecasting of agricultural product futures prices is crucial for farmers, traders and policymakers. (Zhang et al., 2025)

– Predictions enable farmers and businesses to adapt to market changes and safeguard their economic interests. (Zhang et al., 2025)

– Accurate agricultural futures price forecasts can guide the formulation and adjustment of government policies. (Zhang et al., 2025)

– Three main approaches for agricultural product futures price are suggested; traditional forecasting methods, machine learning methods, and combined forecasting models. (Zhang et al., 2025)

Latent Dirichlet Allocation (LDA)

– Topic modeling is a type of natural language processing (NLP) for extracting abstract topics from text data. (Chuluunsaikhan et al., 2020)

– LDA is a tool applied to uncover latent topics for a large corpus. Documents are represented by a mixture of topics, where a topic is a probability distribution over the set of words. (Guindani et al., 2024)

– After the data cleaning, words are extracted according to their parts of speech using KoNLPy. Subsequently, a bag-of-words (BoW) model was created using the result of PoS tagging. The BoW model is the numerical representation of text data, and the output is used in LDA modelling. (Chuluunsaikhan et al., 2020)

– In contrast to BoW-based LDA, Sentence Latent Dirichlet Allocation (Sent-LDA) is proposed that assumes that all words in one sentence are sampled from the same topic. (Li et al., 2022)

– The LDA model returns K number topics, which include M number of words. It is necessary to set K manually, and this is a limitation of the LDA method. (Chuluunsaikhan et al., 2020)

The Sentiment Effect in the Agriculture Futures Market / Constructing a Sentiment Intensity Index Taking into account Investor Concern

– For all agricultural futures investigated, contagious sentiment and idiosyncratic sentiment is significantly and positively associated with futures returns. Contagious sentiment has a considerable impact. Neither domestic nor global stock market sentiment is significantly correlated with futures return (Li et al., 2025).

– Studies stress the use of a 14-day RSI (relative strength index) as a sentiment proxy to construct the sentiment index (Li et al., 2025).

– Sentiment intensity analysis primarily involves the utilization of sentiment-based analysis methods that classify news text using domain-specific sentiment toolkits (An et al., 2024)

– For soybean futures price prediction, the predictive abilities of the proposed sentiment intensity that considers investor concerns are high (An et al., 2024).

– Investor sentiment plays a significant role in influencing trading activities (An et al., 2024).

Food exports boosted by tweets

– The reason for using tweets as a proxy of consumer evaluation is their usefulness for food consumption research. (Inoue et al., 2024)

– It proposes to combine tweets with outcome variables using a time-series analysis. (Inoue et al., 2024)

– A tri-variate vector autoregression (VAR) model was used consisting of net imports of Kiwifruit from New Zealand, unit import price and tweets (Inoue et al., 2024)
– This helped New Zealand to outline its export marketing strategy (Inoue et al., 2024)

Combining Neural Network Prediction Models

– Combining graph convolutional network (GCN) with bidirectional gated recurrent unit (BiGRU) is proposed, and are integrated through tree-structured parzen estimator (TPE). (Zhang et al., 2025)
– Experiments are conducted on a real corn futures dataset, and GCN-BiGRU-TPE shows better prediction performance than traditional machine learning models, single deep learning models, and other hybrid models. (Zhang et al., 2025)
– The neural basis expansion analysis with exogenous variables is improved by designing a weight coefficient and Optuna is used to optimize the designed weight coefficient and the hyperparameters (Wang et al., 2024)

Food Security Themes

Text analysis in Food Security:

– Text classification, a critical part of NLP, aims to automatically categorize textual information into specified taxonomies. Common tasks cover; semantic information retrieval, data filtration, topic modelling, text clustering, Named Entity Recognition (NER) and sentiment analysis (Xiong et al., 2024).
– Text classification models analyze textual data gathered from diverse sources, such as news articles, social media, and online forums.
– Common ML techniques cover; Support Vector Machines (SVMs) and Bayesian classifiers (BC). SVM, grounded in statistical learning theory, excel in intricate classification within high-dimensional feature spaces. Whereas, Bayesian classifiers are less computationally demanding and get good results with limited training sets (Xiong et al., 2024).
– Deep learning-based models provide rich semantic representation of food issues. BERT has demonstrated remarkable potential in text classification and in the food domain, it is used in tasks like food safety supervision and food category text classification. (Xiong et al., 2024).
– The dynamic topic modeling capability of BERTopic extends static text analysis, and allows to observe evolving thematic trends over time (Ma et al., 2025).
– BERTopic model is not only adept at analyzing short texts but also proficient in uncovering the profound semantic structures and temporal evolution of events (Ma et al., 2025).
– News items were filtered that contain at least one keyword from the set of commodity, supply-demand and price-related keywords (Pratap et al., 2022).
– Evolution of Opinion Probabilities serve as a reminder of the importance of managing public emotions and implementing information dissemination strategies in food safety incidents(Ma et al., 2025).

Machine Learning in Food Security:

– ML techniques are playing a crucial role in supporting decision-making in food security applications (Jarray et al., 2023).
– ML techniques in food security cover cropland mapping, crop type mapping, yield forecasting and field delineation (Jarray et al., 2023).
– Several ML techniques were discussed; Support Vector Machines, Random Forest, XGBoost and Kmeans. DL techniques in food security cover; CNN, LSTM, GRU and RNN (Jarray et al., 2023).
– ML application classification, the challenges and future directions, and importance is derived from the fundamental role that food security plays in ensuring a sustainable and equitable future. (Jarray et al., 2023)

Lag in publishing national food price indices/Nowcasting

– Food prices are a complex phenomenon that is influenced by various demand and supply factors. Fluctuations in food prices have the potential to disrupt both international and local market stability (Silva et al., 2024).

– Challenges are posed by the lag in publishing national food price indices and monitoring abnormal growth in food prices worldwide (Silva et al., 2024).

– Nowcasting Food Prices Module provides insights into the current state and trends of food prices at regional, country or territory levels (Silva et al., 2024).

– The proposed nowcasting model utilizes diverse data, ranging from standard variables like the exchange rates from local currency units (LCU) to the US dollar (USD) and the crude oil price, as well as data from non-conventional sources like the crowdsourced daily food prices, and non-conventional variables such as the sentiment index compiled from news articles collected on Twitter (Silva et al., 2024).

Food Inflation

– Food items availability and prices are key factors that affect the nutrition of a population (Silva et al., 2024).

– The informative information content embedded in news data, suggests the use of news-based sentiment indicators as an additional source of information for inflation forecasting (Pratap et al., 2022).

– Modernization of data collection and the use of web scraping techniques (Polidoro et al., 2015)

– Improvements in terms of quality deriving from web scraping for inflation measures are noticed (Polidoro et al., 2015)

Validation of food security indicators

– Rising food prices may rapidly push vulnerable populations into food insecurity (Silva et al., 2024)

– Food security availability and access indicators include several components: quantity (i.e. enough food and energy), quality, safety, and cultural acceptability and preferences. Stability is a cross-cutting dimension that refers to food being available and accessible and utilization being adequate at all times. These measures can be used to make food security statements for groups of individuals and households, as well as at higher levels of aggregation, such as community, national, or global levels (Leroy et al., 2015)

– For food security indicators, ensuring validity is challenging because of the complexity of the construct and the absence of a gold standard measure. Validation exercises should specify first which dimension or component of food security is being assessed and identify the relevant gold standard (Leroy et al., 2015)

– A substantial body of research has been done over past decades to identify, test, and validate indicators of food security access. Future research should build upon this work and focus on harmonizing indicators, setting up a global data collection system to monitor and track progress, and ensuring coordination among actors at all levels; research, practice, and policy (Leroy et al., 2015)

Sentiment analysis using text mining techniques

– In the food domain, sentiment analysis technology is widely applied to outline whether consumers' opinion is directed towards positivity or negativity. Sentiment analysis has progressed from rules-based methods to statistical machine learning and, more recently, deep learning (Xiong et al., 2024).

– Sentiment analysis algorithms are commonly employed on social media platforms to identify and analyze shifts in public sentiment, attitudes, and opinions (Ma et al., 2025).

– Even in the context of relatively stable positive sentiment, sudden spikes in negative sentiment can significantly impact the formation of public opinion. The finding underscores the necessity of real-time monitoring and a sensitive response to emotional dynamics in opinion management and food communication (Ma et al., 2025).

– Coverage of news is outlined and construction of sentiment indicators/indices using text-mining techniques is stressed (Pratap et al., 2022)

Mapping and Defining Vegetation using Machine Learning

– Machine learning can be used to predict crop yields and monitor cropland abandonment (Jarray et al., 2023).

– The remote sensing community has created several procedures for mapping and defining vegetation. They are reliant on environmental elements like soil parameters, hydrography, landform, and meteorological conditions. (Kumaran et al., 2025)

– Vegetation has been thoroughly mapped, monitored, and evaluated both qualitatively as well as quantitatively using spectral indices (Kumaran et al., 2025)

– Data features are extracted and classified using reinforcement radial Gaussian encoder with adversarial Boltzmann temporal neural networks (Kumaran et al., 2025)

In order to extract the buzzwords that characterize textmining integration within food price prediction and food security modelling, we extracted the nouns from the generated titles. Then we calculated the repetitions of each buzzword in the research papers and divided this by the total number of characters of the research papers where they appeared in (Figure3 and Figure4). The normalized repetition score reflects the priority of the buzzword and how it is a cornerstone of the research domain.

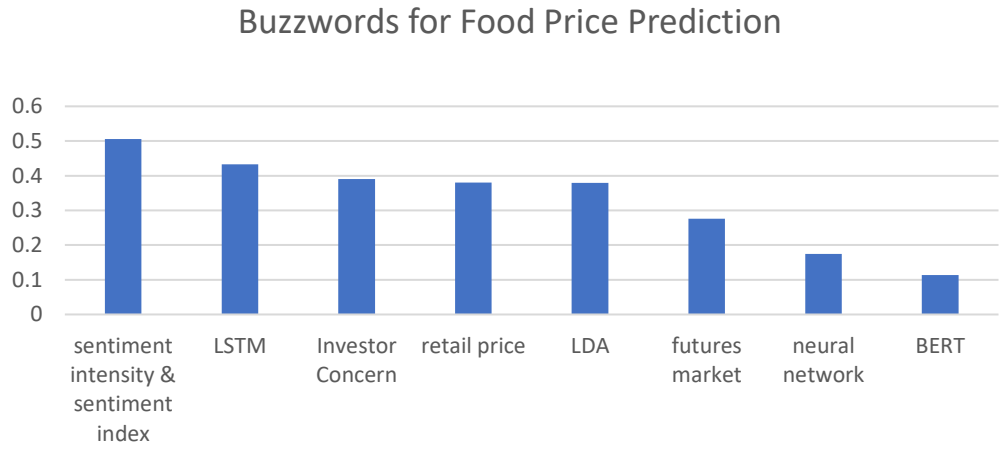

Figure3: Buzzwords for Food Price Prediction

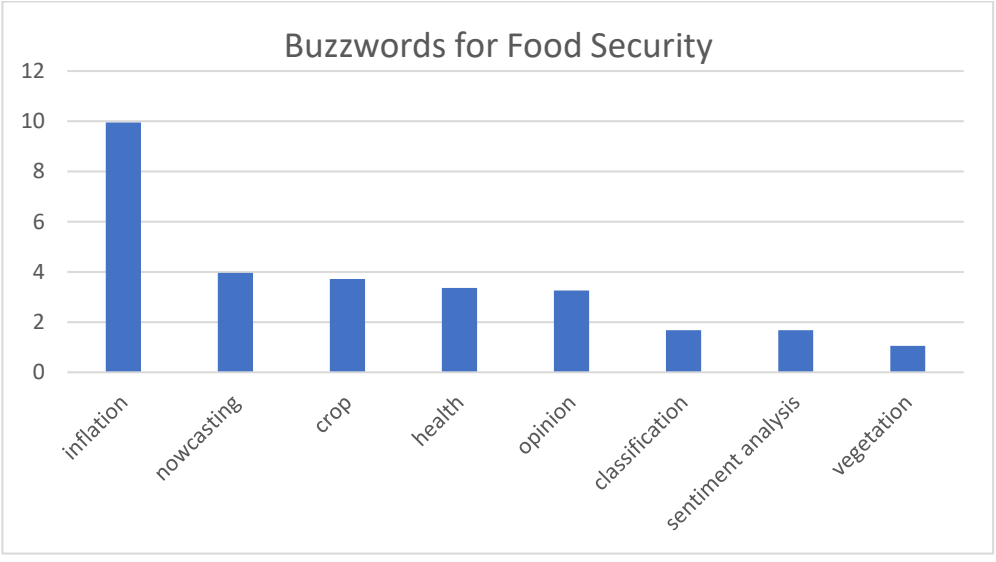

Figure4: Buzzwords for Food Security

Conclusions

The paper aims to outline research themes in the domain of integrating text-based techniques with quantitative food price forecasting and food security models. With the use of SBERT, we provided a semantic representation of the text covered in the research papers. Afterwards, we carried out clustering to achieve a homogeneous breakdown of the content based on similarity. This gave the research directions and concerns tackled in the papers. We conclude that for food price prediction sentiment analysis reflected in sentiment intensity and sentiment indexes play a pivotal role in integrating text-based techniques in quantitative forecasting models. Also, LSTM is found a common, well-proven neural network-based approach for food price prediction. Then comes LDA for topic modeling in food price forecasting, whereas BERT topic models still do not have such wide spread, although BERT has a richer semantic representation. Future work is required to compare bag-of-word LDA and BERT in topic modelling for forecasting, and compare their time performance and accuracy results. In addition, combining neural network models and integrating them with traditional forecasting models has been highlighted. This research direction is worth investigation, and we expect several research in such direction. Tackling investor concern in textminig for food price prediction is stressed in the papers, and we expect it to be a major perspective for the domain.

For food security 'Inflation' was found the major concern, as it has the highest normalized occurrence. This means it is a prime indicator to tackle food security. Therefore, we suggest research to address inflation with the advanced forecasting and text-based techniques. Similarly, nowcasting is a very important research direction in food security, and we suggest future research to address it with state-of-the-art quantitative and text-based techniques. Crop and vegetation have attracted less attention in food security. This can be interpreted as they are concerned more with the food quality rather than the contextual events that can incidentally happen within the supply chain and can cause disruptions. Again, tackling opinion of stakeholders in textmining for food security is stressed. The domain of integrating text-based techniques with quantitative models is still nascent, and with this semantic analysis we aim to conceptualize its constituents and directions.

References

AbdulKader, M., Kumaran, M. S., Keerthika, V., Reddy, P. S., & Rajendra, A. (2024). Security analysis based on sustainable agriculture: Artificial intelligence application. *Remote Sensing in Earth Systems Sciences, 8*(56–64).

An, W., Wang, L., & Zeng, Y.-R. (2024). Social media-based multi-modal ensemble framework for forecasting soybean futures price. *Computers and Electronics in Agriculture, 226*, 109439. https://doi.org/10.1016/j.compag.2024.109439

Chuluunsaikhan, T., Ryu, G.-A., Yoo, K.-H., Rah, H., & Nasridinov, A. (2020). Incorporating deep learning and news topic modeling for forecasting pork prices: The case of South Korea. *Agriculture, 10*(11), 513. https://doi.org/10.3390/agriculture10110513

Guindani, L. G., Oliveira, G. A., Ribeiro, M. H. D. M., Gonzalez, G. V., & de Lima, J. D. (2024). Exploring current trends in agricultural commodities forecasting methods through text mining: Developments in statistical and artificial intelligence methods. *Heliyon, 10*(23), e40568. https://doi.org/10.1016/j.heliyon.2024.e40568

Inoue, Y., & Nakajima, S. (2024). The role of tweets in agricultural export: An approach from text-mining and time-series analyses. *British Food Journal, 126*(4), 1597–1616. https://doi.org/10.1108/BFJ-07-2023-0623

Jarray, N., Ben Abbes, A., & Farah, I. R. (2023). Machine learning for food security: Current status, challenges, and future perspectives. *Artificial Intelligence Review, 56*, 53853–S3876.

Jayadianti, H., Permadi, V., & Partoyo, P. (2023). LSTM forecast of volatile national strategic food commodities. *JURNAL INFOTEL, 15*(4), 345–351.

Kumaran M, M.A., M.S., Keerthika, V. et al. Rural Ecosystem Monitoring in Food Security Analysis Based on Sustainable Agriculture: Artificial Intelligence Application. Remote Sens Earth Syst Sci 8, 56–64 (2025). https://doi.org/10.1007/s41976-024-00166-4

Leroy, J. L., Ruel, M., Frongillo, E. A., Harris, J., & Ballard, T. J. (2015). Measuring the food access dimension of food security: A critical review and mapping of indicators. *Food and Nutrition Bulletin, 36*(2), 167–195.

Li, J., Li, G., Liu, M., Zhu, X., & Wei, L. (2022). A novel text-based framework for forecasting agricultural futures using massive online news headlines. *International Journal of Forecasting, 38*(1), 35–50. https://doi.org/10.1016/j.ijforecast.2020.02.002

Li, Y., Liu, F., & He, W. (2025). Sentiment and futures returns in Chinese agricultural futures markets. *SAGE Open, 15*(2). https://doi.org/10.1177/21582440251335708

Ma, B., & Zheng, R. (2025). Exploring food safety emergency incidents on Sina Weibo: Using text mining and sentiment evolution. *Journal of Food Protection, 88*, 100418.

Polidoro, F., Giannini, R., Lo Conte, R., Mosca, S., & Rossetti, F. (2015). Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation. *Statistical Journal of the IAOS, 31*, 165–176.

Pratap, B., Ranjan, A., Kishore, V., & Bhoi, B. B. (2022). Catching up on the TOP news: Text-mining for food inflation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.XXXXXXX (placeholder DOI)

Reddy, R. K., Jangir, P., Tamilarasi, G., Kumar, R. S., Nitya, E., Bakka, V., & Kumar, G. (2024). Sustainable agriculture-based food security analysis using healthcare data modelling and deep learning techniques. *Remote Sensing in Earth Systems Sciences, 8*(45–55).

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982–3992. Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1410

Silva e Silva, L., Mongeau Ospina, C. A., & Fabi, C. (2024). Food price inflation nowcasting and monitoring. *Statistical Journal of the IAOS, 40*, 325–339.

Vasilescu, C. L., McKee, M., & Reeves, A. (2024). Quantitative textual analysis as a means to explore corporate interests in food safety. *European Commission / LSHTM / LSE*.

Wang, L., An, W., & Li, F.-T. (2024). Text-based corn futures price forecasting using improved neural basis expansion network. *Journal of Forecasting, 43*(6). https://doi.org/10.1002/for.3119

Wang, W., & Liu, Y. (2025). A novel framework for agricultural futures price prediction with BERT-based topic identification and sentiment analysis. *Journal of Forecasting*. https://doi.org/10.1002/for.3278

Wihartiko, F. A., Rahutomo, R., & Budi, I. (2021). Review on artificial intelligence application in agricultural price prediction. *Procedia Computer Science, 179*

Xiong, S., Tian, W., Si, H., Zhang, G., & Shi, L. (2024). A survey of the applications of text mining for the food domain. *Algorithms, 17*(176). https://doi.org/10.3390/a17050176

Xu, L., & Hsu, S. H. (2022). Predicting commodity prices using sentiment analysis and deep learning models. *Journal of Computational Social Science, 5*(3), 765–783. https://doi.org/10.1007/s42001-021-00119-z

Zhang, D., Li, X., Ling, L., et al. (2025). Integrated GCN–BiGRU–TPE agricultural product futures prices prediction based on multi-graph construction. *Computational Economics*. https://doi.org/10.1007/s10614-025-10345-9