algorithms

MDPI

Review

A Survey of the Applications of Text Mining for the Food Domain

Shufeng Xiong, Wenjie Tian, Haiping Si, Guipei Zhang and Lei Shi *

College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002, China; xsf@whu.edu.cn (S.X.); twjtian@stu.henau.edu.cn (W.T.); haiping@henau.edu.cn (H.S.)
* Correspondence: shilei@henau.edu.cn

Abstract: In the food domain, text mining techniques are extensively employed to derive valuable insights from large volumes of text data, facilitating applications such as aiding food recalls, offering personalized recipes, and reinforcing food safety regulation. To provide researchers and practitioners with a comprehensive understanding of the latest technology and application scenarios of text mining in the food domain, the pertinent literature is reviewed and analyzed. Initially, the fundamental concepts, principles, and primary tasks of text mining, encompassing text categorization, sentiment analysis, and entity recognition, are elucidated. Subsequently, an analysis of diverse types of data sources within the food domain and the characteristics of text data mining is conducted, spanning social media, reviews, recipe websites, and food safety reports. Furthermore, the applications of text mining in the food domain are scrutinized from the perspective of various scenarios, including leveraging consumer food reviews and feedback to enhance product quality, providing personalized recipe recommendations based on user preferences and dietary requirements, and employing text mining for food safety and fraud monitoring. Lastly, the opportunities and challenges associated with the adoption of text mining techniques in the food domain are summarized and evaluated. In conclusion, text mining holds considerable potential for application in the food domain, thereby propelling the advancement of the food industry and upholding food safety standards.

Keywords: text mining; food quality control; recipe recommendation; food safety regulation

Citation: Xiong, S.; Tian, W.; Si, H.; Zhang, G.; Shi, L. A Survey of the Applications of Text Mining for the Food Domain. *Algorithms* **2024**, *17*, 176. https://doi.org/10.3390/a17050176

Academic Editors: Mateus Mendes and Balduíno Mateus

Received: 23 March 2024
Revised: 17 April 2024
Accepted: 24 April 2024
Published: 25 April 2024

1. Introduction

Food plays a crucial role in people's daily lives, significantly influencing their health and overall well-being [1]. The textual data related to food mainly come from ingredient lists, nutritional information, and other details found on food packaging. With the continuous advancement of information technology, the amount of accessible textual data in the food domain is rapidly increasing.

Meanwhile, the rapid development of text mining methods [2,3] makes it possible to deeply analyze the rich semantic information contained in the content of the text. It is used in various areas, especially in the food industry, where it plays an important role in analyzing food quality [4], providing gastronomic recommendations [5], warning about food safety issues [6,7], and creating regulations for food security [8], among other things. Specifically, the work in [9] monitored food quality and safety issues based on user-generated content for timely corporate response. The work in [10] utilized big data to provide a comprehensive examination of the entire food supply chain, thereby increasing synergies and efficiencies throughout the supply chain and beyond. The work in [11] performed semantic analysis of food products based on reviews and menu descriptions, providing a way for users to discover their favorite dishes based on personal preferences and descriptions. Recommendation systems [12–14] use text mining techniques to provide guidance on healthy dietary choices, as well as to customize recipes based on individual and group preferences. These examples highlight the potential effectiveness of text mining in providing intelligent decision support in the food domain.

Text mining techniques have the same potential in the food domain. However, food practitioners may not be familiar with the techniques and applications in this domain. This paper aims to fill this knowledge gap by providing a concise introduction to the major text mining techniques and their applications in the food domain. It will also review the latest technological advances in text mining in the food domain to help food practitioners better utilize text mining techniques to solve relevant problems and improve productivity and product quality.

2. Methodology

In the process of conducting this comprehensive review, a thorough literature search was meticulously carried out to gather relevant information from reputable English databases, including the Web of Science, PubMed, IEEE Xplore, Google Scholar, and others, employing strategic keywords using the following terms: 1. "text mining" and "food", 2. "data mining" and "food", 3. "Text Classification" and "food", 4. "Sentiment Analysis" and "food", 5. "Information Extraction" and "food", 6. "Named Entity Recognition" and "food", and 7. "Text Clustering" and "food". This series of searches resulted in an initial set of papers. These papers were screened to exclude those that did not meet the following conditions: 1. duplicates from different indexing repositories, 2. topics other than agriculture and text mining, and 3. a lack of peer review. This set of papers was then further expanded by looking for papers that cited or were cited by this initial batch of papers.

In the literature classification stage, we categorized the selected literature according to different criteria. Firstly, the literature was classified into different categories based on its subject matter or content. Figure 1 visually illustrates the specific application of text mining techniques in the food domain. Secondly, the literature was categorized according to the text mining techniques used; five common techniques were selected for this review, and Figure 2 illustrates the proportion of papers in each area. Finally, we categorized the literature by data source into government and science data, media and social data, and consumer data.

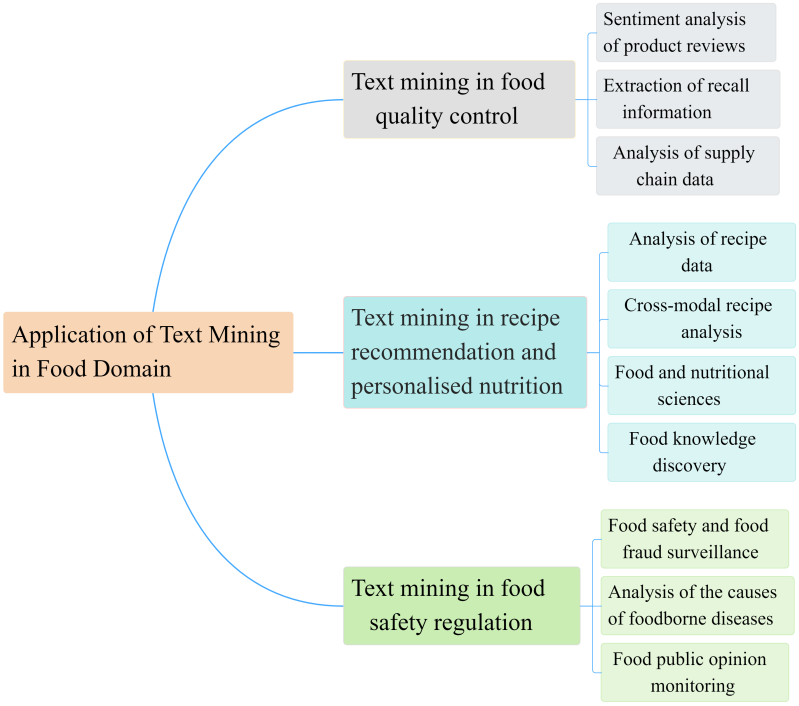

Figure 1. Application classification in the food domain.

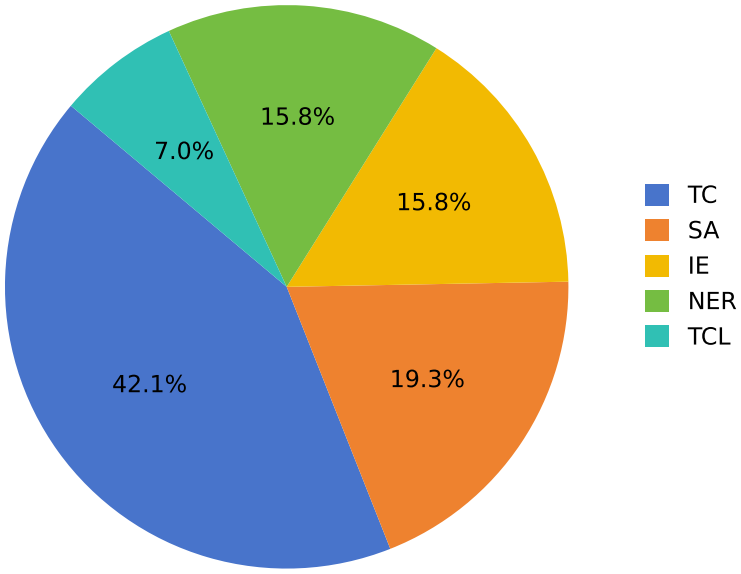

Figure 2. Number of papers, where TC = text classification, SA = sentiment analysis, IE = information extraction, NER = named entity recognition, and TCL = text clustering.

The organizational structure of this review is as follows: text mining overview (Section 3), data sources in the food domain (Section 4), text mining in food quality control (Section 5), text mining in recipe recommendation and personalized nutrition (Section 6), text mining in food safety regulation (Section 7), challenges and future directions (Section 8), and conclusions (Section 9).

3. Text Mining Overview

The rise of text data in society highlights their importance in social interactions. Manual processing was once efficient, but digitization has made it impractical, especially for urgent tasks. Text mining, using natural language processing and machine learning, helps extract insights from large datasets, aiding in understanding text content [3].

The subsequent section will delve into the research scope of the text mining field, with particular emphasis on studies related to the food domain.

3.1. Text Classification

Text classification, a critical part of NLP, is a large endeavor with the main objective of automatically categorizing textual information into specified taxonomies. Its value extends to a wide range of applications across several domains, including spam filtration, news subject categorization, and sentiment analysis [15]. Table 1 lists frequently used text categorization methods, along with their benefits and drawbacks, as well as pertinent studies that use these approaches.

Table 1. General methods of text classification.

Methods	Algorithms	Comments	Papers
Machine learning	KNN, SVM, NB, and so on.	Automated, efficient, generalized, but data-dependent and difficult to interpret.	Monakhova et al. (2011) [16], Wiegand et al. (2013) [17], Chang et al. (2020) [18], Wiegand et al. (2015) [19], Barbara et al. (2018) [20], van den Bulk et al. (2022) [21], Eftimov et al. (2017) [22], Sowinski et al. (2022) [23]

Table 1. *Cont.*

Methods	Algorithms	Comments	Papers
Deep learning	CNN, RNN, Transformer, and so on.	Has contextual understanding advantage, learns rich semantic knowledge, but has high data requirements and computational complexity.	Adyasha et al. (2019) [24], Zuo et al. (2020) [25], Pingzhen et al. (2021) [26], Enguang et al. (2022) [27], Huang et al. (2022) [9], Xiong et al. (2023) [28]

Text classification models within the food domain serve a crucial role in analyzing textual data gathered from diverse sources, such as news articles [28], social media [29], and online forums [30]. They are also useful for identifying product recalls, food safety incidents, and other related occurrences, which makes it easier to promptly notify people about food safety issues [20,24,26,27]. Moreover, a deeper investigation in this field entails interpreting complex associations between food decisions and health state using data from a corpus of natural language texts [17,19].

Machine learning relies on hand-designed features and models, while deep learning learns features and models to represent data through hierarchical learning [16,21,22]. Support Vector Machines (SVMs), grounded in statistical learning theory, excel in intricate classification within high-dimensional feature spaces. For instance, Monakhova et al. [16] utilized a software system employing data analysis methods to automatically identify suspicious food products. In contrast, the Naive Bayes (NB) method relies on the assumption of conditional independence among textual features as it constructs the classification model. For example, Chang et al. [18] leveraged e-invoice big data to craft an automated food safety alert system for the edible oil manufacturing industry.

In recent years, deep learning-based convolutional neural networks and recurrent neural networks have found applications in text classification tasks within the food domain, aiming to acquire intricate semantic feature representations from text and enhancing classification performance [23,25]. Among these approaches, deep learning models like BERT have demonstrated remarkable potential in text classification tasks [9,28]. BERT, a bidirectional Encoder pre-trained language model, accrues substantial linguistic knowledge during pre-training and undergoes fine-tuning for downstream NLP tasks. Research consistently underscores BERT's capacity to improve text classification outcomes [31]. Within the food domain, BERT finds applications in areas such as food safety supervision and food category text classification, consistently surpassing traditional methodologies.

In NLP, text classification is pivotal for assessing various approaches. From traditional machine learning to deep learning, technological advancements drive text categorization and NLP techniques. Text classification's importance lies in its application value and future potential in the food domain.

3.2. Sentiment Analysis

Sentiment analysis, also known as Opinion Mining, is a prominent and extensively researched field within NLP and textual analysis. From the NLP perspective, the core objective of sentiment analysis lies in the extraction of both the subject of the comment and the proclivity of the commenter to convey sentiment towards said subject [15]. In Table 2, commonly used methods for sentiment analysis, along with their advantages and disadvantages, are presented alongside relevant papers that employ these methods.

Table 2. General methods of sentiment analysis.

Methods	Algorithms	Comments	Papers
Dictionary-Based	Sentiment Lexicon Matching Algorithm, Sentiment Intensity Algorithm based on sentiment lexicon, and so on.	Using a sentiment lexicon and calculating based on the sentiment words in the text are simple and effective, but rely on the quality of the sentiment lexicon and are unable to handle complex relationships in the context.	Alamsyah et al. (2015) [32], Ariyasriwatana et al. (2016) [33], Jia (2018) [34]
Rules-Based	Keyword/Phrase Rule Matching Algorithm. and so on.	Interpretable, domain-adaptable, but rule-engineering dependent, difficulty handling complex contexts.	Vidal et al. (2015) [35]
Machine Learning	KNN, SVM, NB, and so on.	Adaptable to handle complex language structures and contexts, but overly dependent on data quality.	Song et al. (2020) [36]
Deep Learning	CNN, RNN, BERT, and so on.	Capable of learning more complex semantic and contextual features, but requires large computational resources and training data.	Hossain et al. (2020) [37]

In the food domain, sentiment analysis technology is widely applied. When assessing entities such as fast-food restaurants, bubble tea shops, and various food products, sentiment analysis technology enables the detailed collection, categorization, and analysis of online consumer feedback and reviews [34]. This analysis not only aims to ascertain whether consumers' evaluations tend towards positivity or negativity, but also, through meticulous research, can pinpoint consumers' specific perspectives on aspects such as taste, texture, packaging, pricing, and other dimensions of food [33].

Sentiment analysis has progressed from rules-based methods to statistical machine learning and, more recently, deep learning. Initially, it relied on dictionaries and rules, then evolved to incorporate machine learning classifiers for improved results. These technologies not only provide businesses with profound insights into consumer preferences, but also directly impact product design and marketing strategies [35,37]. For instance, Alamsyah et al. [32] conducted a rigorous analysis using a vast dataset from Facebook, creating a social network model that offered valuable insights into market dynamics, competitiveness, and segmented markets.

The introduction of neural network models and the utilization of sentiment analysis techniques employing word and sentence vectors enable the autonomous extraction of advanced semantic features, showcasing adaptability across various data domains and language differences. Taking Song et al.'s [36] research as an example, they employed a stacked ensemble learning framework, integrating various base learners such as Naive Bayes, Support Vector Machine, XGBoost, FastText, a convolutional neural network, Long Short-Term Memory, and BERT for evaluating the impact of global food safety news.

Sentiment analysis technology has garnered significant attention in the food industry, progressing from traditional machine learning to the developments in deep learning. This advancement not only propels progress in the food domain, but also stimulates the evolution of related NLP.

3.3. Information Extraction

Information extraction (IE) is a vital and dynamic field of research within NLP. Its main goal is to automatically extract structured information from unstructured text. Information extraction plays a crucial role in identifying entities, events, and relationships between entities in text, converting unstructured textual data into organized information. This pro-

cess provides valuable assistance for knowledge acquisition and text analysis [15]. Table 3 presents commonly used methods for information extraction, outlining their advantages and disadvantages, along with references to papers that employ these methods.

Table 3. General methods of information extraction.

Methods	Algorithms	Comments	Papers
Rules-Based	Pattern Matching Algorithm, Regular Expressions, and so on.	Easily understood and adapted, but difficult to adapt to complex linguistic structures and uncertainty of change.	Wiegand et al. (2012) [38], Wiegand et al. (2012) [39], Wiegand et al. (2014) [40]
Machine Learning	SVM, CRF, HMM, and so on.	Ability to learn models from large amounts of labeled training data, adaptable, but highly dependent on the quality and diversity of training data for model performance.	Wiegand et al. (2012) [41], Gavai et al. (2021) [42]
Deep Learning	CNN, RNN, LSTM, and so on.	Capable of learning more complex semantics and feature representations, but requires many computational resources and labeled data, poor model interpretability.	Cenikj et al. (2021) [43], Zhang et al. (2022) [44], Zuo et al. (2022) [45], Wang et al. (2022) [46]

Information-extraction technology plays a crucial role in the extensive application within the food industry. This technology is utilized to analyze ingredient lists and nutritional labels of food products, extracting key information such as names, carbohydrates, proteins, fats, and vitamins [38,39]. This capability is essential for monitoring food components and understanding the nutritional composition of different food items.

Furthermore, information extraction contributes to real-time monitoring of food safety, enabling regulatory authorities to extract relevant information related to food safety incidents from sources like news reports and social media. This empowers regulatory bodies to swiftly identify potential risks and take timely actions to ensure public health and safety [38,39]. For example, Gavai et al. [42] utilized artificial intelligence to detect novel stimulants in food supplements, employing machine learning to identify 20 new stimulants from the scientific literature and online sources. Through word embeddings and text mining techniques, relevant authorities can proactively discover potential health risks and promptly alert consumers. Additionally, information-extraction technology is applied to establish relationships between food and chemicals by analyzing the biomedical peer-reviewed scientific literature [43].

Information-extraction technology has evolved from basic methods to sophisticated approaches. Initially, information extraction relied on manually crafted rules and templates. In the 1990s, there was a shift towards employing statistical machine learning models, reducing the need for extensive manual labor in feature engineering [41]. In the 21st century, the introduction of deep learning technologies has further propelled the development of information-extraction methods. Various neural network models can autonomously learn semantic features from text, enabling more intelligent and automated information-extraction processes [40,44–46].

In the field of NLP, information-extraction technology has consistently been a key focus in the domain of food. From the evolution of traditional machine learning to the advent of deep learning, this technology has not only driven advancements in information extraction, but has also spurred the evolution of related NLP techniques. Given the crucial applications of information extraction in extracting food-related information and monitoring safety issues, it holds significant value in the food domain.

3.4. Named Entity Recognition

Named entity recognition (NER) is a fundamental task in the field of NLP. Its objective is to identify entities in a text with special significance, such as personal names, geographic locations, and organization names [47]. Entities can be broadly categorized into general and domain-specific types. The key to NER lies in developing models capable of eliminating text ambiguities, accurately delineating entity boundaries, and assigning appropriate categories. Methods for achieving this goal include rules-based approaches, statistical techniques, and deep learning. NER plays a pivotal role in supporting tasks like information extraction, question-answering systems, and relationship extraction. Table 4 summarizes common NER methods along with their advantages and disadvantages, accompanied by relevant literature utilizing these approaches.

Table 4. General methods of named entity recognition.

Methods	Algorithms	Comments	Papers
Rules-Based	Regular Expressions, Lexicon Lookup, and so on.	Easy to understand and implement, but requires writing many rules manually and is difficult to generalize to new domains.	Eftimov et al. (2017) [48], Popovski et al. (2019) [49], Popovski and Kochev et al. (2019) [50]
Deep Learning	CNN, RNN, LSTM, and so on.	Capable of learning more complex semantic and contextual features, but requires large computational resources and labeled data, poor model interpretability.	Yadav et al. (2018) [51], Li et al. (2020) [47], Cenikj et al. (2022) [52], Perera et al. (2022) [53], Makridis et al. (2023) [54]

The application of NER has greatly expanded in the field of food, encompassing the identification of entities such as ingredient names, brand names, and product specifications in food menus, packaging, and review texts. This provides possibilities for various downstream applications [51]. For example, Eftimov et al. [48] employed a rules-based NER method named drNER to extract knowledge on evidence-based dietary advice from unstructured text.

In NER tasks within the food domain, it is necessary to identify domain-specific terms such as ingredient names, dishes, and food technologies. Additionally, NER models must handle specific language conventions, such as ingredient omissions and pronoun usage in menus. Establishing an NER system for the food domain requires the construction of domain-specific texts to create annotated datasets and the incorporation of contextual features and vocabulary constraints in the text to enhance the model's recognition capabilities. For instance, Popovski et al. [49] proposed a rules-based NER method called FoodIE for extracting entities related to food from unstructured text data. This method integrates rules from computational linguistics and semantic information to describe features of food entities, addressing shortcomings in information-extraction methods specifically designed for food concepts in NLP. In further research, FoodIE and a gold standard corpus named FoodBase were utilized for an in-depth exploration of the food domain [50]. Additionally, researchers introduced a scientific food NER and Named Entity Linking (NEL) model called SciFoodNER, which fine-tunes a Transformer model and leverages a scientific abstract corpus annotated with food entities [52].

The evolution of NER technology has traversed from initial rules-based methodologies to statistical learning techniques and, notably, to the era of deep learning. Within deep learning, various neural network models, including CNN, RNN, LSTM, and BERT, have substantially enhanced NER's performance [47]. For instance, Perera et al. [53] compared the performance of seven text mining models for food and dietary ingredient NER, revealing differences between classical machine learning models and the latest deep language models. Contemporary NER systems can be trained end-to-end, eliminating the need for manual feature engineering, thanks to pre-trained language models. Additionally,

the integration of knowledge graphs and transfer learning has improved the adaptability of NER systems [54].

NER stands as a cornerstone technology within NLP, with significant application potential in the food domain. Its evolution from rules-based to deep learning methodologies has not only enhanced performance, but also broadened its application scope. As language technology and artificial intelligence continue to advance, NER is expected to play an increasingly pivotal role in supporting intelligent decision-making and services.

3.5. Text Clustering

Text clustering is an unsupervised machine learning paradigm that autonomously categorizes unstructured textual data based on inherent similarities [15]. Its primary utility lies in its ability to reveal underlying themes and concepts within textual corpora, condensing extensive datasets by identifying textual resemblances and grouping similar compositions into coherent clusters. Unlike supervised learning, text clustering does not require manual data labeling, making it particularly effective for managing large textual archives. Common methodologies include distance-based clustering techniques, such as hierarchical clustering algorithms, partitioning algorithms, and hybrid strategies that combine hierarchical and partitioned clustering approaches [15]. Table 5 presents commonly used methods for text clustering, outlining their advantages and disadvantages, along with references to papers that employ these methods.

Table 5. General methods of text clustering.

Methods	Algorithms	Comments	Papers
Hierarchical Clustering	KNN, Complete-Linkage Clustering, and so on.	No need to pre-specify the number of clusters, able to capture hierarchical structure, but higher computational complexity for large-scale datasets, more sensitive to outliers and noise.	Singh et al. (2018) [55], Pigłowski et al. (2019) [56]
Non-Hierarchical Clustering	K-Means Clustering, Grid-Based Clustering, and so on.	Computationally efficient for irregularly shaped clusters, but sensitive to initial values.	Lee et al. (2013) [57], Kim et al. (2018) [58]

Text clustering has gained significant traction within the food domain, particularly in endeavors such as the extraction and categorization of themes from consumer feedback. This application enhances the ability to analyze and predict consumer preferences and tendencies related to food, utilizing social network-based consumer ratings and feedback [58]. An interesting application involves the use of a fuzzy relational clustering algorithm to classify Korean food items based on the adjectives used to describe their flavors. This approach resulted in the establishment of linguistic scales for various gastronomic categories, benefiting consumers in selecting their preferred foods and providing valuable insights to purveyors and producers about consumer requisites and inclinations [57]. Leveraging text-clustering technology, food enterprises gain the capability to extract meaningful insights from textual datasets, enlightening them about consumer needs and the dynamics of public sentiment [56].

As an unsupervised text analytical modality, text clustering serves as a pivotal instrument for the classification and scrutiny of extensive unstructured textual data. Its intersection with the food landscape continues to grow and evolve alongside advances in machine learning and natural language processing. Consequently, it provides the alimentary industry with an indispensable tool for consumer demand analysis and public opinion monitoring, emerging as a guiding star in the arena of data analytics [55].

4. Data Sources in the Food Domain

Textual data in the food domain come from diverse sources such as government agencies, corporations, media outlets, and consumers, offering insights into food safety, distribution, and consumer trends [3]. Government and corporate datasets provide quantitative information, while media and social data offer qualitative perspectives on food health, quality, and trends. Consumer-generated content like product reviews reflects consumer sentiments. This amalgamation of data provides a comprehensive view of the food domain, essential for decision making and research.

In this section, we delve into the utilization of these three primary information streams, elucidating their unique contributions and practical applications in gastronomic text mining. To facilitate easy access and reference, we augmented Table 6 with detailed descriptions of each source and direct links to relevant resources.

Table 6. Literature dataset sources.

Author	Public or Not	Source
F.J. van de Brug et al. (2014) [59]	No	Government and science data
Kathryn Montgomery et al. (2017) [60]	No	Government and science data
Yamine Bouzembrak et al. (2016) [61] [1]	Yes	Government and science data
Alberto Nogales et al. (2020) [62]	No	Government and science data
Anand Gavai et al. (2023)[63]	No	Government and science data
Lee et al. (2023) [64]	No	Government and science data
M. Pigłowski et al. (2019) [56]	No	Government and science data
Shanquan Chen et al. (2016) [56]	No	Government and science data
Shweta Singh Chauhan et al. (2020) [56] [2]	Yes	Government and science data
Kasper Jensen et al. (2015)[65] [3]	Yes	Government and science data
Wahiba Ben Abdessalem Karaa et al. (2018)[66] [4]	Yes	Government and science data
Hui Yang et al. (2011) [67]	No	Government and science data
Ruth Areli García-León et al. (2019)[68]	No	Media and social data
Sofiane Abbar et al. (2015)[1]	No	Media and social data
Debarchana (Debs) Ghosh et al. (2013)[69]	No	Media and social data
Haoyang Zhang et al. (2023)[6] [5]	Yes	Media and social data
Daniel Fried et al. (2014)[70] [6]	Yes	Media and social data
Mohamed M. Mostafa et al. (2018)[29]	No	Media and social data
Yuru Huang et al. (2019)[71]	No	Media and social data
Munmun De Choudhury et al. (2016)[72]	No	Media and social data
Kate G. Blackburn et al. (2018)[73]	No	Media and social data
R.Akila et al. (2020)[4]	No	Consumer data
Shuting Tao et al. (2022)[74]	No	Consumer data
Leticia Vidal et al. (2015) [72]	No	Consumer data
Andry Alamsyah et al. (2015)[35]	No	Consumer data
Adyasha Maharana et al. (2019) [24]	No	Consumer data

[1] https://www.foodchainid.com/products/food-fraud-database (accessed on 17 April 2024). [2] http://ctf.iitrindia.org/focusdb/ (accessed on 17 April 2024). [3] http://cbs.dtu.dk/services/NutriChem-1.0 (accessed on 17 April 2024). [4] https://pubmed.ncbi.nlm.nih.gov/(accessed on 17 April 2024). [5] https://github.com/DachuanZhang-FutureFood/IFoodCloud (accessed on 17 April 2024). [6] https://sites.google.com/site/twitter4food/ (accessed on 17 April 2024).

There are a wide range of sources of textual data in the food domain, but there are a number of challenges and limitations. Despite the wide range of data sources, not all of the data are publicly available. Many of these datasets may be subject to confidentiality restrictions or proprietary in nature, which limits further use of the data by researchers and analysts.

4.1. Government and Science Data

Governmental and scientific data play a prominent role in the food domain, derived from statistical reports, assay records, surveillance bulletins, and related publications from

governmental bodies. This corpus of information provides a comprehensive macro-level view of the state of affairs in the entire food domain, overseeing the quality and safety of food and shaping food policies [59,60,63,75]. For example, scholars have utilized data from the European Rapid Early Warning System for Food and Feed (RASFF) to predict food deception and anticipate risks to food safety [61–64]. Notification data on mycotoxins from the RASFF database have been employed for statistical analyses to decipher the risk associated with mycotoxins in food [56].

Governmental databases house repositories of national food safety standards, results from food sampling, monitoring of foodborne pathogens, and licenses for food production. These databases aggregate food standards, assay results, and surveillance data, serving as foundational infrastructure for comprehensive monitoring and risk assessment of food quality and safety. They provide crucial information support for the decision-making processes of food regulators [76]. For instance, the creation of a food safety information database within the Greater China region empowers governmental entities with the capacity to analyze and compare food safety, facilitating the formulation of effective policies [76]. Computational toxicologists at the Indian Institute of Toxicology created FOCUS-DB, a comprehensive repository compiling detailed information on 2885 food additives, serving as a valuable tool for scientific inquiry and public awareness [77].

Scientific literature exploration is another source of textual information, where researchers leverage scientific abstracts to glean insights into associations among plant-based foods, phytochemical constituents, and human diseases. For instance, "NutriChem", a scholarly resource, connects the chemical makeup of plant-based foods with the phenotypes of human diseases, providing a foundation for well-informed nutritional and therapeutic strategies [65]. Scientific literature research is crucial for scrutinizing relationships between foods, genetics, and diseases [66,67], fostering a deep understanding of these intricate relationships.

It is crucial to emphasize that repositories sourced from governmental agencies, international bodies, and the scientific realm adhere to rigorous data integrity standards, enhancing their data trustworthiness. Most publicly available datasets are also derived from government and science data.

4.2. Media and Social Data

In recent years, media and social data have played an increasingly pivotal role in the information ecosystem of the food domain. This type of data primarily comes from public news outlets, informational releases by industry organizations, and data exchanges among consumers across various social platforms [68]. Online social platforms actively generate real-time textual data, enabling the analysis of behaviors, dietary patterns, and health monitoring [1,68,69].

Media data not only monitor and report significant occurrences, policy trajectories, and industrial advancements in the food domain, but also provide timely insights into public concerns, highlighting hotspots and focal points. They serve as an indispensable source for surveilling food-related public opinions, supporting research, and facilitating decision-making processes [6]. Systematic collation and content analysis of media reports allow the identification of evolving trends and issues, enabling the anticipation of public sentiment trajectories and empowering relevant authorities to respond swiftly and accurately. Concurrently, media data provide valuable insights into the industry. Social data, on the other hand, present an abundance of fragmented, user-generated content with expansive coverage and a profusion of information. For example, by analyzing the linguistic patterns of food-related conversations on social media and their correlations with demographic characteristics, researchers have uncovered intricate associations between food-related discourse and geographical, as well as community attributes [70].

The comprehensive use of data from various social platforms enables access to a diverse cross-section of individuals, capturing genuine user sentiments. This equips enterprises with insights to enhance user experience and product optimization and enables

regulatory bodies to align themselves with the voice of consumers [4,29]. In comparison to media data, social data encompass more subjective and emotionally infused content. However, the unstructured nature of such data increases the complexity of analysis and data mining. For instance, Huang et al. [71] used geographically tagged food-related tweets on Twitter to analyze food environments and chronic health outcomes across distinct U.S. census regions. By harnessing data from social media to analyze the uneven distribution of sustenance and resources, as well as the motivations, attitudes, beliefs, and emotions expressed by individuals in their online discourse concerning food, researchers discuss the instrumental role of social media in ameliorating disparities in food accessibility and health [72,73].

Currently, media and social data are extremely time-sensitive, so there is an urgent need to develop standardized tools and intelligent analytical models, while at the same time digging deeper into the intrinsic value of textual data. This has become an important development direction for contemporary research.

4.3. Consumer Data

Consumer data, as a vital source of textual information in the food domain, offer insights into the authentic experiences and perceptions of consumers. Extracting and analyzing this reservoir of data is crucial for organizations to align with the genuine voices of their users and formulate strategies focused on user experiences [4,74,78]. These data primarily come from diverse sources, including restaurant reviews, e-commerce product evaluations, social media platforms, and various forms of user-generated content. Social data, in particular, reflect individual sentiments and emotions, providing valuable insights for understanding consumer behavior and experiences in various dietary contexts [35].

It is important to note that different platforms attract distinct user cohorts. By amalgamating data from various sources, a more comprehensive user profile can be crafted. The textual analysis of consumer data is foundational for product enhancement and informed marketing decision making [32]. Beyond production and marketing, consumer data analyses support regulatory efforts in the food domain. User feedback serves as an early indicator for regulatory responses, correlating issues raised by consumers with subsequent food recall notifications and incidents. This expedites food safety measures, mitigates the impact on public health and the economy, and ensures the timely identification of unsafe food products [24].

Current research still faces significant challenges in dealing with the integration and intelligent analysis of massive amounts of consumer data. Among them, advancing natural language processing techniques for deeper understanding of unstructured textual data has become one of the key issues to address. At the same time, the protection of user privacy becomes the most urgent and important consideration when utilizing such data. Consumer data are generally collected and labeled by individuals for scientific research. However, as most of the datasets are not publicly available, they cannot be used for further research.

5. Text Mining in Food Quality Control

5.1. Sentiment Analysis of Product Reviews

Product reviews are valuable collections of user opinions, reflecting personal perspectives and consumer experiences. The use of NLP techniques for sentiment analysis, theme mining, and relationship extraction of food reviews, combined with comprehensive analysis of user data, enables the creation of consumer profiles and personalized quality control. For example, Yong et al. [79] proposed an innovative model that combines sentiment analysis with BERT, a pre-trained text model, to enhance the understanding of online food reviews.

However, these analyses face challenges, such as the complexity and diversity of consumer language and the need to verify review authenticity. Researchers are actively improving text representation and sentiment understanding to uncover subtle quality concerns. Meaningful categorization, as demonstrated by Ariyasriwatana et al. [33], can

contribute to healthier eating practices and assist policymakers and food companies in developing effective programs and products. work, empowers developers and marketers to better understand consumer needs and preferences, ultimately boosting the quality and sales of food products.

In the effort to analyze what consumers say about the taste, smell, and overall characteristics of food in online discussions, one challenge arises from the lack of specialized words related to the senses. Kim et al. [58] used a technique called word embedding, based on skip-word modeling, to study how people express their opinions about food on social media. This method predicts whether consumers will like a food without relying on specialized sensory words or expensive lab tests. By adopting this approach, food developers and marketers can better understand what consumers want, improving the quality and sales of food products. Similarly, Lee et al. [57] introduced a method based on rough set theory to analyze clusters of Korean food and the adjectives people use to describe its taste. This method helps reveal the patterns and characteristics of flavors in Korean cuisine, making it easier for individuals to understand and describe the taste of Korean food. It also provides a new tool for researching and promoting Korean cuisine.

In summary, examining and understanding consumer reviews are crucial for achieving user-focused quality control in the food industry. They are often used in text classification and sentiment analysis techniques, thus allowing food companies to really listen to the voice of the customer, thus potentially transforming and advancing quality management philosophies and approaches. This not only opens the door to utilizing the capabilities of emerging technology, but also represents a vital path for the food industry to enhance its quality management capabilities [30].

5.2. Extraction of Recall Information

Food recalls directly impact consumer well-being and safety. The timely extraction of recall information is crucial for food enterprises to proactively initiate recalls and prevent hazards from escalating. Researchers are increasingly exploring text mining to analyze unstructured data sources, such as recall notices and media reports, aiming to enhance recall responses and strengthen food quality and safety. For example, Maharana et al. [24] utilized text matching techniques to correlate food reviews on Amazon.com with FDA recalls between 2012 and 2014. This innovative approach leveraged consumer reviews to identify unsafe food products promptly, mitigating their impact on public health and the economy. It also addressed challenges related to urbanization, underreporting of illnesses, and tracing the connection between tainted food and subsequent illnesses.

Contemporary research focuses on analyzing unstructured data with a limited text size and complex semantic expressions. Researchers work on improving information-extraction techniques and integrating this information with structured databases for precise recall delineation. Makridis et al. [80] proposed a deep learning and machine learning approach, using time-series prediction and historical recall announcements to foresee future recalls by type. This proactive approach enables timely recalls, enhancing food safety across the supply chain. The use of data-augmentation methodologies further expands the depth and breadth of data sources. Deep learning for predicting food recalls enhances overall safety and control of the food system, contributing to the development of an intelligent and collaborative food regulatory framework [54].

Mining and analyzing recall notices are essential tools for food companies to conduct swift and effective recalls. Researchers using techniques such as text categorization and named entity recognition were able to extract key information from a large number of recall notices, helping food companies accurately identify the affected products, the cause of the recall, and the affected regions so that they can take targeted action.

5.3. Analysis of Supply Chain Data

Traceability is critical in the food supply chain for rapidly identifying the origins of quality issues [81]. Text mining technology offers a means to explore traceability data

at each node, detecting vulnerabilities and risk control points across the supply chain. Researchers are increasingly integrating text analytics to enhance the informational value of traceability systems, contributing to ongoing improvements in quality management throughout food enterprises [82].

A significant challenge in the food supply chain is food waste, estimated at about one-third of the total production [83]. The Internet of Things (IoT) and blockchain technology are emerging solutions, providing transparency, sustainability, and efficiency by collecting data at every stage of the supply chain [84,85].

Text mining tools empower food companies to identify crucial control nodes where product quality may be compromised. Extracting business intelligence from consumer opinions can optimize food production. For example, a case study using text mining on Twitter posts revealed key issues with beef products, offering insights for developing a consumer-centric beef supply chain [55]. Digital technologies are also enhancing the efficiency of the retail food supply chain, transforming the entire system to be more sustainable [82].

Mining and analyzing traceability text data are essential for improving the quality control of the food supply chain. By analyzing traceability text data using information extraction, as well as text clustering techniques, food companies can gain a better understanding of the product's production process, the nodes in the supply chain, and related transaction and shipping information. The deep integration of text mining with emerging technologies like the IoT and blockchain will further drive the digital transformation of the food supply chain, ushering in an era of increased efficiency and transparency.

6. Text Mining in Recipe Recommendation and Personalized Nutrition

6.1. Analysis of Recipe Data

The importance of recipe data lies in the use of text mining methods, which help to analyze huge food datasets and extract structured insights from them to support personalized recipe suggestions and nutritional strategies. Using NLP, essential details such as ingredients, measurements, and procedures are extracted from unstructured recipe narratives and transformed into organized datasets. These foundational data facilitate the exploration of ingredient relationships, assessment of substitutes, and understanding the synergistic properties of ingredient combinations [86]. This analytical capability enables the customization of recipe recommendations based on users' gastronomic preferences and dietary restrictions, allowing for transformations like turning a calorie-dense Western dish into a nutrition-conscious, Asian-inspired variant [87].

Additionally, the analysis of recipe content provides insights into nutritional profiles, allowing assessments of caloric content, macronutrient distributions, and other nutritional parameters [88]. Machine learning applications, such as predicting nutritional values based on textual descriptors, are invaluable for individuals managing specific health conditions, optimizing athletic performance, or pursuing wellness and fitness goals [89].

On the other hand, the meticulous examination of consumer feedback on recipes aids in creating precise models of user food preferences. Applying association rule mining algorithms to historical food interactions helps deduce individualized taste preferences, while scrutinizing online food platforms provides a macroscopic view of dietary preferences and emerging food trends [90–92].

As the global demand for animal-based proteins puts a strain on the environment, the transition towards plant-centric diets becomes crucial for sustainability. Researchers have explored online recipe databases to understand dietary customs, offering insights into regional gastronomic heritage [93]. Additionally, methodologies like block-based linked data generation facilitate the communal exchange and discovery of recipe information, transcending the food domain for broader applicability [94].

To sum up, text mining is a revolutionary method of evaluating large recipe collections and is essential to customized nutrition services and intelligent meal recommendation systems. The field of food and nutritional services is projected to witness a substantial

growth in the use of text mining due to the increased emphasis on health consciousness and the digitization of the food industry.

6.2. Cross-Modal Recipe Analysis

Cross-modal food analysis involves synthesizing textual, pictorial, and multi-modal datasets to capture the essence of recipes and intricacies of ingredient preparation, offering a deeper understanding of cultural and health narratives. By dissecting the visual and procedural attributes of food, such as color and slicing techniques, this approach enables finely tuned, visually oriented food recommendations [95].

Cross-modal methodologies can deduce unarticulated aspects of recipes. Studies, like the work by Chen et al. [96], leverage imagery and textual content to correlate recipes with visual depictions, guiding users in the food-creation process based on visual cues from food photographs. By combining text and image data from diverse sources, a more nuanced system for recipe comprehension and recommendation can be engineered. Innovations, like the R2GAN model introduced by Zhu et al. [97], generate food images from recipe descriptions, enhancing cross-modal recipe retrieval. This advancement benefits individuals seeking to identify dishes from recipes or recipes from images, opening possibilities for cross-modal retrieval in different domains. Cross-modal recipe retrieval, matching food images with recipe scripts or vice versa, has been explored in various studies [98–102]. Looking ahead, as multimodal interaction evolves, additional modalities like auditory and gestural inputs may be integrated, providing users with more intelligent and bespoke food selection and guidance services.

In conclusion, cross-modal analysis improves recipe understanding and marks a major advancement towards customized services and intelligent food recommendation systems by overcoming the limits of text-only data.

6.3. Food and Nutritional Sciences

In the contemporary era, global health concerns are increasingly linked to diet-related chronic diseases resulting from deviations from nutritional norms. Scholarly efforts are underway to establish correlations between individuals' nutritional profiles and overall wellness [89]. The field of food science and nutrition has generated a vast amount of textual data, both structured and unstructured, ready for analysis through advanced text mining methodologies to extract key insights on dietary and nutritional knowledge. Recent research emphasizes the adoption of a quantitative index reflecting food processing levels and physicochemical attributes, moving away from traditional nutrient-centric approaches [103].

Text mining is instrumental in examining documentary sources such as dietary regulations and nutrition facts labels, automating the extraction of essential knowledge on the interrelation among food, nutrients, and diseases. Pioneering work by Yang et al. [104] introduced the Nutritional Epidemiology Ontology (ONE), a framework based on normative guidelines, enriching comprehensive knowledge graphs for intelligent querying and decision support. Text mining also serves as a valuable tool for synthesizing and distilling nutrition research literature, providing scholars with expedited access to scientific discoveries, enhancing research productivity. Additionally, text mining aids in scrutinizing nutrient composition studies, as exemplified by do Nascimento et al. [105], who employed text mining to analyze gluten-free versus gluten-laden provisions. This investigation highlights the compositional nuances and constraints of gluten-free products, emphasizing the need for individuals with celiac disease to be vigilant about nutritional content. Aiello et al. [106] used digital datasets to elucidate Londoners' dietary patterns and their link to metabolic syndrome-associated pathologies, revealing counterintuitive findings.

There is much promise in text mining technology for analyzing scientific data related to nutrition and food. Its capabilities hold the potential to spur new discoveries, improve the effectiveness of research, and create more intelligent knowledge-based services, all of which will advance empirical illumination in the fields of food science and nutritional practices.

6.4. Food Knowledge Discovery

Techniques for data mining are used in the field of knowledge discovery in modern food science to uncover underlying patterns and guiding principles from massive datasets. This method not only contributes to a deeper knowledge of topics like customer preferences, food safety, nutrition, and supply chain logistics, but it also establishes the framework for extensive and thorough study.

Ontologies, structured frameworks for categorizing domain-specific concepts, have become indispensable tools in this era of the Semantic Web [107]. Various ontologies, such as FoodWiki, AGROVOC, Open Food Facts, Food Product Ontology, and Foodon, serve distinct purposes in elucidating the complexities of the food domain [108,109]. For example, FoodWiki analyzes the adverse effects of food additives using Semantic Web technologies to provide tailored advice on safe consumption for different risk demographics [110,111]. Foodon addresses challenges in disparate datasets within food safety, quality, production, distribution, and consumption, aiming to enhance traceability and transparency [109]. The ISO-FOOD ontology encapsulates metadata and traceability elements, supporting future analytical endeavors and integrating stable isotope data within food science research [107]. It seamlessly integrates with pre-existing ontologies like the Unit Metrics Ontology, the Food Nutrition Ontology, and the Literature Ontology, showcasing the synergy of shared knowledge in this domain. The Nutritional Epidemiology Ontology (ONE) standardizes the outputs of nutritional epidemiology studies, embedding data standards, reporting guidelines, and core concepts from authoritative guidelines within the field [104]. An ontology-based repository for food lexicon and methodologies further enhances the exchange and retrieval of food data [94].

Text mining, as a tool for extracting latent knowledge from unstructured textual data, is poised to catalyze the evolution of food science research and knowledge-based services.

7. Text Mining in Food Safety Regulation

7.1. Food Safety and Food Fraud Surveillance

The issue of food fraud, encompassing intentional substitution, adulteration, dilution, falsification, or mislabeling of food products for economic gain, has become a focal point of regulatory attention [112,113]. This illicit practice not only undermines consumer confidence, but also poses serious risks to food safety and quality. Establishing a robust food-safety-monitoring infrastructure for comprehensive surveillance of the entire supply chain is crucial.

Textual data play a crucial role in supporting food safety regulatory mechanisms [114]. Publicly accessible articles, government databases, academic research repositories, and Internet sources contain vast amounts of unstructured textual information. Researchers have compiled databases and utilized data from various sources to analyze patterns and trends in food fraud, predict incidents, and proactively monitor food safety issues [61,115–118]. Social media platforms have also emerged as valuable data reservoirs for monitoring food safety issues globally [119]. Systems like FoodSIS in Singapore exemplify the use of Internet data retrieval for proactive monitoring [120]. Studies on the role of social media in communication disparities related to food safety incidents further underscore its significance [121].

The increasing complexity and globalization of the food supply chain have heightened concerns about food safety, posing a significant threat to consumer health [122]. NLP and text mining are deployed for intelligent textual analysis to foster proactive monitoring and risk alerting for food safety events and fraudulent practices [123,124]. The application of these technologies in analyzing online media data for public health safeguarding and crisis anticipation is advocated [125]. Food safety regulators face a constant influx of reports and complaints in various formats, and the application of multi-class classification techniques aids in monitoring and preventing violations, strengthening consumer rights, public health, and socio-economic stability [126,127].

In the realm of social media, insights from user-generated content are valuable for understanding public perception and sentiment regarding food safety [128,129]. The

convergence of text classification, sentiment analysis, and web-monitoring technologies proves instrumental in shaping discourse on food safety, enabling regulators and the public to navigate this complex domain more effectively.

7.2. Analysis of the Causes of Foodborne Diseases

The surge in foodborne illnesses, caused by contamination with microbial, viral, or parasitic agents, poses a serious threat to public health and safety, impacting both individual well-being and the broader healthcare and economic landscape. Extensive textual data serve as a valuable resource for understanding the causes behind these diseases. Thakur et al. [130] have pioneered a novel data mining methodology to uncover patterns in the incidence of foodborne diseases. Using a combination of attribute selection, decision tree training, and association rule generation, they revealed important correlations among different categories of foodborne afflictions, implicated food sources, and consumption locations.

Additionally, emerging sources of information, such as social media, have been leveraged to enhance the detection of foodborne disease outbreaks [119]. Sadilek et al. [131] developed nEmesis, a system that intelligently identifies potential public health threats from food establishments by analyzing Twitter discourse. This showcases the potential of social media analytics in public health and provides actionable intelligence to strengthen public health responses. Similarly, collaborations between health departments and academic institutions, like the one between the New York City Department of Health and Mental Hygiene and Columbia University [132], have utilized online restaurant reviews on platforms such as Yelp to identify previously unreported outbreaks of foodborne illnesses, offering valuable insights into the origin and impact of these diseases [133,134].

Text data analysis of foodborne illness has challenges such fragmentation, compartmentalization, and quality inconsistencies. The development of extensive text analysis tools and the establishment of defined methods for text annotation are crucial in addressing these difficulties. Hu et al. [135] presented TWEET-FID, an annotated dataset designed for the multifaceted detection of foodborne illness events. They provided a detailed overview of the robust framework for creating this dataset and the fine-grained annotation process and showed how state-of-the-art deep learning methods can be applied to various tasks using the TWEET-FID corpus.

In conclusion, a comprehensive and effective analysis of foodborne disease etiologies necessitates the integration of various methodologies. Text data offer new perspectives for exploration through the use of text classification, information extraction, and sentiment analysis, and continued efforts to advance text intelligence analytics are critical to ensuring food safety.

7.3. Food Public Opinion Monitoring

In the face of ongoing food security challenges, monitoring public sentiment on food safety is essential for understanding incident impacts and crafting targeted regulatory approaches. This surveillance hinges on vast unstructured data from modern media channels. Kate et al. [136], for example, applied text mining to extract instances of public grievances about food safety from various web forums, enhancing governmental capacity to compile a comprehensive database of food safety concerns and improving regulatory oversight precision.

Insights gained from monitoring public perspectives on food enable regulators to quickly identify and address focal points of public discourse, proactively mitigate societal distress, and navigate collective sentiment with targeted precision. For example, the IFoodCloud platform, offering real-time analysis of China's food safety public opinion, aggregates data from over 3100 sources. This platform helps decipher public sentiment, analyze regional disparities, and understand public concerns following food safety incidents [137]. In the aftermath of a food safety crisis, regulators can trigger responsive protocols informed by monitoring outputs to disseminate accurate information and maintain societal equilib-

rium. Zhang et al. [6] showcased the power of big data and machine learning in analyzing public sentiment around food safety in Greater China, particularly during the initial stages of the 2019 NKP outbreak, highlighting the potential of these technologies in enhancing risk communication and decision making.

In essence, food opinion monitoring is an essential tool for enacting precise and informed supervisory actions. Food safety regulators continue to raise the bar on food safety governance by utilizing techniques such as text classification, sentiment analysis, etc., to enhance effectiveness and enlightenment.

8. Challenges and Future Directions

8.1. Opportunities

In the food domain, text mining technology offers promising applications, especially in food safety monitoring, business value analysis, and consumer support, with ample room for further development.

Text mining plays a crucial role in supporting food safety regulations by analyzing data from social media, online forums, and government sources. This enables early detection of foodborne disease outbreaks and helps identify problematic foods, providing decision-making support to regulatory authorities [130–132,134]. Additionally, text analysis can monitor dietary patterns and population health, allowing for targeted dietary guidance and health education [1,69,138].

In the business landscape, text mining provides food enterprises with valuable market intelligence. By analyzing consumer feedback on digital platforms, businesses can understand customer preferences, informing product innovation and marketing strategies [33,58].

Consumers also benefit from text mining technology. Through the analysis of diverse data such as purchase records and browsing behavior, mobile applications and recommendation systems can offer personalized dietary advice and shopping guidance, facilitating informed food choices and promoting healthy eating habits [111,139,140].

8.2. Challenges

The food domain possesses a vast amount of unstructured text data, including recipes, menus, and reviews, which are valuable assets. Text mining involves preprocessing and feature extraction, followed by machine learning and deep learning algorithms for tasks like classification, clustering, and sentiment analysis. However, challenges arise due to the varied quality, diversity, and complexity of multi-source text data in the food domain. Ensuring semantic understanding is hindered by inconsistent quality, including misspellings and grammar errors in reviews. Authenticating reviews and dealing with potential false information further complicate the analysis. Additionally, the lack of structured metadata limits semantic understanding, despite efforts to generate structured training data through manual annotation. Data diversity and complexity necessitate effective methods for integration, cleaning, and management, especially concerning different languages, formats, and sources [141]. Ethical and legal concerns, such as privacy protection, also arise, balancing open application with information security [142]. Overall, while text mining offers promising applications in the food industry, addressing challenges of data quality, representativeness, completeness, privacy, and skill requirements is essential for its effective implementation and social value creation.

9. Conclusions

This paper provides a comprehensive survey of text mining techniques in the food domain and their potential impact on industry development. It delineates the effective utilization of text mining for extracting valuable insights and outlines its applications in food recalls, personalized recipes, and food safety regulation. Furthermore, it presents a comprehensive overview of text mining in the food domain, encompassing basic concepts, data source analysis, food quality control, personalized recipe recommendation, and food safety monitoring. Finally, it highlights the opportunities and challenges associated with

text mining technology in the food industry, emphasizing its pivotal role in industry advancement and the establishment of food safety standards.

The utilization of text data as a pivotal communication and information conduit in human society is well established. The integration of text mining technology with artificial intelligence, big data, and other state-of-the-art methodologies to construct a text-based intelligent decision support system holds significant promise for elevating the sophistication of food processing, quality control, and food services. It is essential to acknowledge the inherent trade-off between data quality and personal privacy protection in big data analysis, necessitating further research to achieve an optimal equilibrium. As text mining technology advances further, the latent value embedded within textual data is gradually being unearthed and leveraged, thereby perpetuating the advancement of the entire food industry.

Author Contributions: Conceptualization, S.X.; data curation, W.T.; methodology, L.S.; resources, G.Z. and H.S.; supervision, S.X.; writing—original draft, W.T.; writing—review and editing, H.S., L.S. and S.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received support from Henan Province key research and development project (No. 231111211300).

Data Availability Statement: No data were used for the research described in the article.

Conflicts of Interest: All authors declare no conflicts of interest.

References

1. Abbar, S.; Mejova, Y.; Weber, I. You tweet what you eat: Studying food consumption through twitter. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Republic of Korea, 18–23 April 2015; pp. 3197–3206.
2. Hobbs, J.R.; Walker, D.E.; Amsler, R.A. Natural language access to structured text. In Proceedings of the Coling 1982: Proceedings of the Ninth International Conference on Computational Linguistics, Prague, Czech Republic, 5–10 July 1982.
3. Zhai, C.; Massung, S. *Text Data Management and Analysis: A Practical Introduction to Information Retrieval and Text Mining*; Morgan & Claypool: San Rafael, CA, USA, 2016.
4. Akila, R.; Revathi, S.; Shreedevi, G. Opinion mining on food services using topic modeling and machine learning algorithms. In Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 6–7 March 2020; pp. 1071–1076.
5. Qin, L.; Hao, Z.; Zhao, L. Food safety knowledge graph and question answering system. In Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City, Shanghai, China, 20–23 December 2019; pp. 559–564.
6. Zhang, H.; Zhang, D.; Wei, Z.; Li, Y.; Wu, S.; Mao, Z.; He, C.; Ma, H.; Zeng, X.; Xie, X.; et al. Analysis of public opinion on food safety in Greater China with big data and machine learning. *Curr. Res. Food Sci.* **2023**, *6*, 100468. [CrossRef] [PubMed]
7. Xiao, K.; Wang, C.; Zhang, Q.; Qian, Z. Food safety event detection based on multi-feature fusion. *Symmetry* **2019**, *11*, 1222. [CrossRef]
8. Steen, B.; Marvin, H. Development of food fraud media monitoring system based on text mining. *Food Control* **2018**, *93*, 283–296.
9. Huang, Y.; Wang, X.; Wang, R.; Min, J. Analysis and recognition of food safety problems in online ordering based on reviews text mining. *Wirel. Commun. Mob. Comput.* **2022**, *2022*, 4209732. [CrossRef]
10. Ding, X.; Xu, S. Food safety pre-warning system based on Robust Principal Component Analysis and Improved Apriori Algorithm. In Proceedings of the 2021 International Conference on Pattern Recognition and Intelligent Systems,Bangkok, Thailand, 28–30 July 2021; pp. 5–9.
11. Keshtkar, F.; Shi, L.; Bukhari, S.A.C. The Semantics and Collocations Relation in Food Reviews. In Proceedings of the International FLAIRS Conference Proceedings, Virtually Event, 16–19 May 2021; Volume 34.
12. Ge, M.; Elahi, M.; Fernaández-Tobías, I.; Ricci, F.; Massimo, D. Using tags and latent factors in a food recommender system. In Proceedings of the 5th International Conference on Digital Health 2015, Florence, Italy, 18–20 May 2015; pp. 105–112.
13. Trang Tran, T.N.; Atas, M.; Felfernig, A.; Stettinger, M. An overview of recommender systems in the healthy food domain. *J. Intell. Inf. Syst.* **2018**, *50*, 501–526. [CrossRef]
14. Haussmann, S.; Seneviratne, O.; Chen, Y.; Ne'eman, Y.; Codella, J.; Chen, C.H.; McGuinness, D.L.; Zaki, M.J. FoodKG: A semantics-driven knowledge graph for food recommendation. In Proceedings of the Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, 26–30 October 2019; Proceedings, Part II 18; Springer: Berlin/Heidelberg, Germany, 2019; pp. 146–162.
15. Aggarwal, C.C.; Aggarwal, C.C. *Mining Text Data*; Springer: Berlin/Heidelberg, Germany, 2015.
16. Monakhova, Y.B.; Löbell-Behrends, S.; Maixner, S.; Böse, W.; Marx, G.; Lachenmeier, D.W. Automated classification of web pages for identification of suspicious food products-a feasibility study. *Dtsch. Lebensm. Rundsch.* **2011**, *107*, 328–330.

17. Wiegand, M.; Klakow, D. Towards the detection of reliable food-health relationships. In Proceedings of the Workshop on Language Analysis in Social Media, Atlanta, GA, USA, 13 June 2013; pp. 69–79.
18. Chang, W.T.; Yeh, Y.P.; Wu, H.Y.; Lin, Y.F.; Dinh, T.S.; Lian, I.b. An automated alarm system for food safety by using electronic invoices. *PLoS ONE* **2020**, *15*, e0228035. [CrossRef] [PubMed]
19. Wiegand, M.; Klakow, D. Detecting conditional healthiness of food items from natural language text. *Lang. Resour. Eval.* **2015**, *49*, 777–830. [CrossRef]
20. Koroušić Seljak, B.; Korošec, P.; Eftimov, T.; Ocke, M.; Van der Laan, J.; Roe, M.; Berry, R.; Crispim, S.P.; Turrini, A.; Krems, C.; et al. Identification of requirements for computer-supported matching of food consumption data with food composition data. *Nutrients* **2018**, *10*, 433. [CrossRef] [PubMed]
21. van den Bulk, L.M.; Bouzembrak, Y.; Gavai, A.; Liu, N.; van den Heuvel, L.J.; Marvin, H.J. Automatic classification of literature in systematic reviews on food safety using machine learning. *Curr. Res. Food Sci.* **2022**, *5*, 84–95. [CrossRef] [PubMed]
22. Eftimov, T.; Korošec, P.; Koroušić Seljak, B. StandFood: Standardization of foods using a semi-automatic system for classifying and describing foods according to FoodEx2. *Nutrients* **2017**, *9*, 542. [CrossRef] [PubMed]
23. Sowinski, P.; Wasielewska-Michniewska, K.; Ganzha, M.; Paprzycki, M. Topical Classification of Food Safety Publications with a Knowledge Base. In *Sustainable Technology and Advanced Computing in Electrical Engineering*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 673–693.
24. Maharana, A.; Cai, K.; Hellerstein, J.; Hswen, Y.; Munsell, M.; Staneva, V.; Verma, M.; Vint, C.; Wijaya, D.; Nsoesie, E.O. Detecting reports of unsafe foods in consumer product reviews. *JAMIA Open* **2019**, *2*, 330–338. [CrossRef] [PubMed]
25. Zuo, M.; He, S.Y.; Zhang, Q.C.; Wang, Q.B. Character-word Double-dimensional Semantic Classification Model for Judging Illegal and Irregular Behaviors for Internet Food Safety. In Proceedings of the 2020 IEEE 20th International Conference on Software Quality, Reliability and Security Companion (QRS-C), Macau, China, 11–14 December 2020; pp. 571–577.
26. Wu, P.; Wu, W.; Yuan, S. Research on consumers' perception of food risk based on LSTM sentiment classification. *Food Sci. Technol.* **2021**, *42*, e47221.
27. Zuo, E.; Aysa, A.; Muhammat, M.; Zhao, Y.; Chen, B.; Ubul, K. A food safety prescreening method with domain-specific information using online reviews. *J. Consum. Prot. Food Saf.* **2022**, *17*, 163–175. [CrossRef]
28. Xiong, S.; Tian, W.; Batra, V.; Fan, X.; Xi, L.; Liu, H.; Liu, L. Food safety news events classification via a hierarchical transformer model. *Heliyon* **2023**, *9*, e17806. [CrossRef] [PubMed]
29. Mostafa, M.M. Mining and mapping halal food consumers: A geo-located Twitter opinion polarity analysis. *J. Food Prod. Mark.* **2018**, *24*, 858–879. [CrossRef]
30. Gan, Q.; Ferns, B.H.; Yu, Y.; Jin, L. A text mining and multidimensional sentiment analysis of online restaurant reviews. *J. Qual. Assur. Hosp. Tour.* **2017**, *18*, 465–492. [CrossRef]
31. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186.
32. Alamsyah, A.; Peranginangin, Y. Network market analysis using large scale social network conversation of Indonesia's fast food industry. In Proceedings of the 2015 3rd International Conference on Information and Communication Technology (ICoICT), Nusa Dua, Bali, Indonesia, 27–29 May 2015; pp. 327–331.
33. Ariyasriwatana, W.; Quiroga, L.M. A thousand ways to say'Delicious!'—Categorizing expressions of deliciousness from restaurant reviews on the social network site Yelp. *Appetite* **2016**, *104*, 18–32. [CrossRef]
34. Jia, S. Behind the ratings: Text mining of restaurant customers' online reviews. *Int. J. Mark. Res.* **2018**, *60*, 561–572. [CrossRef]
35. Vidal, L.; Ares, G.; Machín, L.; Jaeger, S.R. Using Twitter data for food-related consumer research: A case study on "what people say when tweeting about different eating situations". *Food Qual. Prefer.* **2015**, *45*, 58–69. [CrossRef]
36. Song, B.; Shang, K.; He, J.; Yan, W.; Zhang, T. Impact assessment of food safety news using stacking ensemble learning. In *Transdisciplinary Engineering for Complex Socio-Technical Systems–Real-Life Applications*; IOS Press: Amsterdam, The Netherlands, 2020; pp. 353–362.
37. Hossain, N.; Bhuiyan, M.R.; Tumpa, Z.N.; Hossain, S.A. Sentiment analysis of restaurant reviews using combined CNN-LSTM. In Proceedings of the 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 1–3 July 2020; pp. 1–5.
38. Wiegand, M.; Roth, B.; Klakow, D. Web-based relation extraction for the food domain. In Proceedings of the Natural Language Processing and Information Systems: 17th International Conference on Applications of Natural Language to Information Systems, NLDB 2012, Groningen, The Netherlands, 26–28 June 2012; Proceedings 17; Springer: Berlin/Heidelberg, Germany, 2012; pp. 222–227.
39. Wiegand, M.; Roth, B.; Lasarcyk, E.; Köser, S.; Klakow, D. A gold standard for relation extraction in the food domain. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12), Istanbul, Turkey, 21–27 May 2012; European Language Resources Association: Paris, France, 2012; pp. 507–514.
40. Wiegand, M.; Roth, B.; Klakow, D. Automatic food categorization from large unlabeled corpora and its impact on relation extraction. In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, 26–30 April 2014; pp. 673–682.

41. Wiegand, M.; Roth, B.; Klakow, D. Data-driven Knowledge Extraction for the Food Domain. In Proceedings of the 11th Conference on Natural Language Processing (KONVENS 2012), Empirical Methods in Natural Language Processing, Vienna, Austria, 19–21 September 2012; Österreichische Gesellschaft für Artificial Intelligence: Vienna, Austria, 2012; pp. 21–29.
42. Gavai, A.K.; Bouzembrak, Y.; van den Bulk, L.M.; Liu, N.; van Overbeeke, L.F.; van den Heuvel, L.J.; Mol, H.; Marvin, H.J. Artificial intelligence to detect unknown stimulants from scientific literature and media reports. *Food Control* **2021**, *130*, 108360. [CrossRef]
43. Cenikj, G.; Seljak, B.K.; Eftimov, T. FoodChem: A food-chemical relation extraction model. In Proceedings of the 2021 IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, USA, 5–7 December 2021; pp. 1–8.
44. Zhang, Q.; Li, M.; Dong, W.; Zuo, M.; Wei, S.; Song, S.; Ai, D. An Entity Relationship Extraction Model Based on BERT-BLSTM-CRF for Food Safety Domain. *Comput. Intell. Neurosci.* **2022**, *2022*, 7773259. [CrossRef] [PubMed]
45. Zuo, M.; Zhang, B.; Zhang, Q.; Yan, W.; Ai, D. An Entity Relation Extraction Method for Few-Shot Learning on the Food Health and Safety Domain. *Comput. Intell. Neurosci.* **2022**, *2022*, 1879483. [CrossRef] [PubMed]
46. Wang, Q.; Zhang, Q.; Zuo, M.; He, S.; Zhang, B. A entity relation extraction model with enhanced position attention in food domain. *Neural Process. Lett.* **2022**, *54*, 1449–1464. [CrossRef]
47. Li, J.; Sun, A.; Han, J.; Li, C. A survey on deep learning for named entity recognition. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 50–70. [CrossRef]
48. Eftimov, T.; Koroušić Seljak, B.; Korošec, P. A rule-based named-entity recognition method for knowledge extraction of evidence-based dietary recommendations. *PLoS ONE* **2017**, *12*, e0179488. [CrossRef] [PubMed]
49. Popovski, G.; Kochev, S.; Eftimov, T. FoodIE: A Rule-based Named-entity Recognition Method for Food Information Extraction. In Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2019), Prague, Czech Republic, 19–21 February 2019.
50. Popovski, G.; Seljak, B.K.; Eftimov, T. FoodBase corpus: A new resource of annotated food entities. *Database* **2019**, *2019*, baz121. [CrossRef]
51. Yadav, V.; Bethard, S. A Survey on Recent Advances in Named Entity Recognition from Deep Learning models. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, NM, USA, 20–26 August 2018; pp. 2145–2158.
52. Cenikj, G.; Petelin, G.; Seljak, B.K.; Eftimov, T. SciFoodNER: Food Named Entity Recognition for Scientific Text. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; pp. 4065–4073.
53. Perera, N.; Nguyen, T.T.L.; Dehmer, M.; Emmert-Streib, F. Comparison of text mining models for food and dietary constituent named-entity recognition. *Mach. Learn. Knowl. Extr.* **2022**, *4*, 254–275. [CrossRef]
54. Makridis, G.; Mavrepis, P.; Kyriazis, D. A deep learning approach using natural language processing and time-series forecasting towards enhanced food safety. *Mach. Learn.* **2023**, *112*, 1287–1313. [CrossRef]
55. Singh, A.; Shukla, N.; Mishra, N. Social media data analytics to improve supply chain management in food industries. *Transp. Res. Part E Logist. Transp. Rev.* **2018**, *114*, 398–415. [CrossRef]
56. Pigłowski, M. Comparative analysis of notifications regarding mycotoxins in the Rapid Alert System for Food and Feed (RASFF). *Qual. Assur. Saf. Crop. Foods* **2019**, *11*, 725–735. [CrossRef]
57. Lee, J.; Ghimire, D.; Rho, J.O. Rough clustering of Korean foods based on adjectives for taste evaluation. In Proceedings of the 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Shenyang, China, 23–25 July 2013; pp. 472–475.
58. Kim, A.Y.; Ha, J.G.; Choi, H.; Moon, H. Automated text analysis based on skip-gram model for food evaluation in predicting consumer acceptance. *Comput. Intell. Neurosci.* **2018**, *2018*, 9293437. [CrossRef] [PubMed]
59. Van de Brug, F.; Luijckx, N.L.; Cnossen, H.; Houben, G. Early signals for emerging food safety risks: From past cases to future identification. *Food Control* **2014**, *39*, 75–86. [CrossRef]
60. Montgomery, K.; Chester, J.; Nixon, L.; Levy, L.; Dorfman, L. Big Data and the transformation of food and beverage marketing: Undermining efforts to reduce obesity? *Crit. Public Health* **2019**, *29*, 110–117. [CrossRef]
61. Bouzembrak, Y.; Marvin, H.J. Prediction of food fraud type using data from Rapid Alert System for Food and Feed (RASFF) and Bayesian network modelling. *Food Control* **2016**, *61*, 180–187. [CrossRef]
62. Nogales, A.; Morón, R.D.; García-Tejedor, Á.J. Food safety risk prediction with Deep Learning models using categorical embeddings on European Union data. *arXiv* **2020**, arXiv:2009.06704.
63. Gavai, A.; Bouzembrak, Y.; Mu, W.; Martin, F.; Kaliyaperumal, R.; van Soest, J.; Choudhury, A.; Heringa, J.; Dekker, A.; Marvin, H.J. Applying federated learning to combat food fraud in food supply chains. *npj Sci. Food* **2023**, *7*, 46. [CrossRef] [PubMed]
64. Lee, N.Z.H. Named Entity Extraction for Food Safety Events Monitoring. Master's Thesis, Nanyang Technological University, Singapore, 2023.
65. Jensen, K.; Panagiotou, G.; Kouskoumvekaki, I. NutriChem: A systems chemical biology resource to explore the medicinal value of plant-based foods. *Nucleic Acids Res.* **2015**, *43*, D940–D945. [CrossRef] [PubMed]
66. Ben Abdessalem Karaa, W.; Mannai, M.; Dey, N.; Ashour, A.S.; Olariu, I. Gene-disease-food relation extraction from biomedical database. In Proceedings of the Soft Computing Applications: Proceedings of the 7th International Workshop Soft Computing Applications (SOFA 2016); Springer: Berlin/Heidelberg, Germany, 2018; Volume 17; pp. 394–407.

67. Yang, H.; Swaminathan, R.; Sharma, A.; Ketkar, V.; D 'Silva, J. Mining biomedical text towards building a quantitative food-disease-gene network. In *Learning Structure and Schemas from Documents*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 205–225.
68. Areli García-León, R. Twitter and food well-being: analysis of# SlowFood postings reflecting the food well-being of consumers. *Glob. Media J. México* **2019**, *16*, 5.
69. Ghosh, D.; Guha, R. What are we 'tweeting'about obesity? Mapping tweets with topic modeling and Geographic Information System. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 90–102. [CrossRef] [PubMed]
70. Fried, D.; Surdeanu, M.; Kobourov, S.; Hingle, M.; Bell, D. Analyzing the language of food on social media. In Proceedings of the 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, USA, 27–30 October 2014; pp. 778–783.
71. Huang, Y.; Huang, D.; Nguyen, Q.C. Census tract food tweets and chronic disease outcomes in the US, 2015–2018. *Int. J. Environ. Res. Public Health* **2019**, *16*, 975. [CrossRef] [PubMed]
72. De Choudhury, M.; Sharma, S.; Kiciman, E. Characterizing dietary choices, nutrition, and language in food deserts via social media. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco, CA, USA, 27 February–2 March 2016; pp. 1157–1170.
73. Blackburn, K.G.; Yilmaz, G.; Boyd, R.L. Food for thought: Exploring how people think and talk about food online. *Appetite* **2018**, *123*, 390–401. [CrossRef] [PubMed]
74. Tao, S.; Kim, H.S. Online customer reviews: Insights from the coffee shops industry and the moderating effect of business types. *Tour. Rev.* **2022**, *77*, 1349–1364. [CrossRef]
75. Lucas Luijckx, N.B.; van de Brug, F.J.; Leeman, W.R.; van der Vossen, J.M.; Cnossen, H.J. Testing a text mining tool for emerging risk identification. *EFSA Support. Publ.* **2016**, *13*, 1154E. [CrossRef]
76. Chen, S.; Huang, D.; Nong, W.; Kwan, H.S. Development of a food safety information database for Greater China. *Food Control* **2016**, *65*, 54–62. [CrossRef]
77. Chauhan, S.S.; Sachan, D.K.; Parthasarathi, R. FOCUS-DB: An Online Comprehensive Database on Food Additive Safety. *J. Chem. Inf. Model.* **2020**, *61*, 202–210. [CrossRef] [PubMed]
78. Marvin, H.J.; Janssen, E.M.; Bouzembrak, Y.; Hendriksen, P.J.; Staats, M. Big data in food safety: An overview. *Crit. Rev. Food Sci. Nutr.* **2017**, *57*, 2286–2295. [CrossRef] [PubMed]
79. Yong, L.; Yang, X.; Liu, Y.; Liu, R.; Jin, Q. A new emotion analysis fusion and complementary model based on online food reviews. *Comput. Electr. Eng.* **2022**, *98*, 107679. [CrossRef]
80. Makridis, G.; Mavrepis, P.; Kyriazis, D.; Polychronou, I.; Kaloudis, S. Enhanced food safety through deep learning for food recalls prediction. In Proceedings of the Discovery Science: 23rd International Conference, DS 2020, Thessaloniki, Greece, 19–21 October 2020, Proceedings 23; Springer: Berlin/Heidelberg, Germany, 2020; pp. 566–580.
81. Badia-Melis, R.; Mishra, P.; Ruiz-García, L. Food traceability: New trends and recent advances. A review. *Food Control* **2015**, *57*, 393–401. [CrossRef]
82. El Bilali, H.; Allahyari, M.S. Transition towards sustainability in agriculture and food systems: Role of information and communication technologies. *Inf. Process. Agric.* **2018**, *5*, 456–464. [CrossRef]
83. Rejeb, A.; Keogh, J.G.; Treiblmaier, H. Leveraging the internet of things and blockchain technology in supply chain management. *Future Internet* **2019**, *11*, 161. [CrossRef]
84. Astill, J.; Dara, R.A.; Campbell, M.; Farber, J.M.; Fraser, E.D.; Sharif, S.; Yada, R.Y. Transparency in food supply chains: A review of enabling technology solutions. *Trends Food Sci. Technol.* **2019**, *91*, 240–247. [CrossRef]
85. Kamilaris, A.; Fonts, A.; Prenafeta-Boldύ, F.X. The rise of blockchain technology in agriculture and food supply chains. *Trends Food Sci. Technol.* **2019**, *91*, 640–652. [CrossRef]
86. Pellegrini, C.; Özsoy, E.; Wintergerst, M.; Groh, G. Exploiting Food Embeddings for Ingredient Substitution. In Proceedings of the HEALTHINF, Virtual Conference, 11–13 February 2021; pp. 67–77.
87. Morales-Garzón, A.; Gómez-Romero, J.; Martin-Bautista, M.J. Semantic-aware transformation of short texts using word embeddings: An application in the Food Computing domain. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop, Virtual Conference, 19–23 April 2021; pp. 148–154.
88. Ispirova, G.; Eftimov, T.; Seljak, B.K. Predefined domain specific embeddings of food concepts and recipes: A case study on heterogeneous recipe datasets. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; pp. 4074–4083.
89. Ispirova, G.; Eftimov, T.; Seljak, B.K. Exploring Knowledge Domain Bias on a Prediction Task for Food and Nutrition Data. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 3563–3572.
90. Ueda, M.; Takahata, M.; Nakajima, S. User's food preference extraction for personalized cooking recipe recommendation. In Proceedings of the Workshop of ISWC, Bonn, Germany, 23–24 October 2011; pp. 98–105.
91. Asano, Y.M.; Biermann, G. Rising adoption and retention of meat-free diets in online recipe data. *Nat. Sustain.* **2019**, *2*, 621–627. [CrossRef]
92. Komariah, K.S.; Sin, B.K. Enhancing Food Ingredient Named-Entity Recognition with Recurrent Network-Based Ensemble (RNE) Model. *Appl. Sci.* **2022**, *12*, 10310. [CrossRef]

93. Rong, C.; Liu, Z.; Huo, N.; Sun, H. Exploring Chinese dietary habits using recipes extracted from websites. *IEEE Access* **2019**, *7*, 24354–24361. [CrossRef]
94. Öztürk, Ö.; Özacar, T. A case study for block-based linked data generation: Recipes as jigsaw puzzles. *J. Inf. Sci.* **2020**, *46*, 419–433. [CrossRef]
95. Chen, J.j.; Ngo, C.W.; Chua, T.S. Cross-modal recipe retrieval with rich food attributes. In Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; pp. 1771–1779.
96. Chen, J.; Pang, L.; Ngo, C.W. Cross-modal recipe retrieval: How to cook this dish? In Proceedings of the MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavik, Iceland, 4–6 January 2017; Proceedings, Part I 23; Springer: Berlin/Heidelberg, Germany, 2017; pp. 588–600.
97. Zhu, B.; Ngo, C.W.; Chen, J.; Hao, Y. R2gan: Cross-modal recipe retrieval with generative adversarial network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11477–11486.
98. Pham, H.X.; Guerrero, R.; Pavlovic, V.; Li, J. CHEF: Cross-modal hierarchical embeddings for food domain retrieval. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual Conference, 2–9 February 2021; Volume 35, pp. 2423–2430.
99. Zhu, B.; Ngo, C.W.; Chen, J.j. Cross-domain cross-modal food transfer. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 3762–3770.
100. Chen, J.J.; Ngo, C.W.; Feng, F.L.; Chua, T.S. Deep understanding of cooking procedure for cross-modal recipe retrieval. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea, 22–26 October 2018; pp. 1020–1028.
101. Wang, H.; Sahoo, D.; Liu, C.; Lim, E.p.; Hoi, S.C. Learning cross-modal embeddings with adversarial networks for cooking recipes and food images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11572–11581.
102. Salvador, A.; Gundogdu, E.; Bazzani, L.; Donoser, M. Revamping cross-modal recipe retrieval with hierarchical transformers and self-supervised learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference, 19–25 June 2021; pp. 15475–15484.
103. Fardet, A.; Lakhssassi, S.; Briffaz, A. Beyond nutrient-based food indices: A data mining approach to search for a quantitative holistic index reflecting the degree of food processing and including physicochemical properties. *Food Funct.* **2018**, *9*, 561–572. [CrossRef]
104. Yang, C.; Ambayo, H.; De Baets, B.; Kolsteren, P.; Thanintorn, N.; Hawwash, D.; Bouwman, J.; Bronselaer, A.; Pattyn, F.; Lachat, C. An ontology to standardize research output of nutritional epidemiology: From paper-based standards to linked content. *Nutrients* **2019**, *11*, 1300. [CrossRef]
105. do Nascimento, A.B.; Fiates, G.M.R.; Dos Anjos, A.; Teixeira, E. Analysis of ingredient lists of commercially available gluten-free and gluten-containing food products using the text mining technique. *Int. J. Food Sci. Nutr.* **2013**, *64*, 217–222. [CrossRef]
106. Aiello, L.M.; Schifanella, R.; Quercia, D.; Del Prete, L. Large-scale and high-resolution analysis of food purchases and health outcomes. *EPJ Data Sci.* **2019**, *8*, 14. [CrossRef]
107. Eftimov, T.; Ispirova, G.; Potočnik, D.; Ogrinc, N.; Seljak, B.K. ISO-FOOD ontology: A formal representation of the knowledge within the domain of isotopes for food science. *Food Chem.* **2019**, *277*, 382–390. [CrossRef]
108. Kamel Boulos, M.N.; Yassine, A.; Shirmohammadi, S.; Namahoot, C.S.; Brückner, M. Towards an "Internet of Food": food ontologies for the internet of things. *Future Internet* **2015**, *7*, 372–392. [CrossRef]
109. Dooley, D.M.; Griffiths, E.J.; Gosal, G.S.; Buttigieg, P.L.; Hoehndorf, R.; Lange, M.C.; Schriml, L.M.; Brinkman, F.S.; Hsiao, W.W. FoodOn: A harmonized food ontology to increase global food traceability, quality control and data integration. *npj Sci. Food* **2018**, *2*, 23. [CrossRef]
110. Çelik, D. Foodwiki: Ontology-driven mobile safe food consumption system. *Sci. World J.* **2015**, *2015*, 475410. [CrossRef]
111. Celik Ertuğrul, D. FoodWiki: A mobile app examines side effects of food additives via semantic web. *J. Med. Syst.* **2016**, *40*, 41. [CrossRef]
112. Spink, J.; Moyer, D.C. Defining the public health threat of food fraud. *J. Food Sci.* **2011**, *76*, R157–R163. [CrossRef]
113. Marvin, H.J.; Bouzembrak, Y.; Janssen, E.M.; van der Fels-Klerx, H.V.; van Asselt, E.D.; Kleter, G.A. A holistic approach to food safety risks: Food fraud as an example. *Food Res. Int.* **2016**, *89*, 463–470. [CrossRef]
114. Fritsche, J. Recent developments and digital perspectives in food safety and authenticity. *J. Agric. Food Chem.* **2018**, *66*, 7562–7567. [CrossRef]
115. Moore, J.C.; Spink, J.; Lipp, M. Development and application of a database of food ingredient fraud and economically motivated adulteration from 1980 to 2010. *J. Food Sci.* **2012**, *77*, R118–R126. [CrossRef]
116. Bouzembrak, Y.; Camenzuli, L.; Janssen, E.; Van der Fels-Klerx, H. Application of Bayesian Networks in the development of herbs and spices sampling monitoring system. *Food Control* **2018**, *83*, 38–44. [CrossRef]
117. Bouzembrak, Y.; Marvin, H.J. Impact of drivers of change, including climatic factors, on the occurrence of chemical food safety hazards in fruits and vegetables: A Bayesian Network approach. *Food Control* **2019**, *97*, 67–76. [CrossRef]
118. Yang, Y.; Wei, L.; Pei, J. Application of Bayesian modelling to assess food quality & safety status and identify risky food in China market. *Food Control* **2019**, *100*, 111–116.
119. Meyer, C.; Hamer, M.; Terlau, W.; Raithel, J.; Pongratz, P. Web data mining and social media analysis for better communication in food safety crises. *Int. J. Food Syst. Dyn.* **2015**, *6*, 129–138.

120. Kate, K.; Chaudhari, S.; Prapanca, A.; Kalagnanam, J. FoodSIS: A text mining system to improve the state of food safety in singapore. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 1709–1718.
121. Zhu, X.; Huang, I.Y.; Manning, L. The role of media reporting in food safety governance in China: A dairy case study. *Food Control* **2019**, *96*, 165–179. [CrossRef]
122. King, T.; Cole, M.; Farber, J.M.; Eisenbrand, G.; Zabaras, D.; Fox, E.M.; Hill, J.P. Food safety for food security: Relationship between global megatrends and developments in food safety. *Trends Food Sci. Technol.* **2017**, *68*, 160–175. [CrossRef]
123. Liu, N.; Bouzembrak, Y.; Van den Bulk, L.M.; Gavai, A.; van den Heuvel, L.J.; Marvin, H.J. Automated food safety early warning system in the dairy supply chain using machine learning. *Food Control* **2022**, *136*, 108872. [CrossRef]
124. Goldberg, D.M.; Khan, S.; Zaman, N.; Gruss, R.J.; Abrahams, A.S. Text mining approaches for postmarket food safety surveillance using online media. *Risk Anal.* **2022**, *42*, 1749–1768. [CrossRef] [PubMed]
125. Zhang, J.; Chen, M.; Hu, E.; Wu, L. Data mining model for food safety incidents based on structural analysis and semantic similarity. *J. Ambient. Intell. Humaniz. Comput.* **2020**, 1–15. [CrossRef]
126. Magalhães, G.; Faria, B.M.; Reis, L.P.; Cardoso, H.L. Text mining applications to facilitate economic and food safety law enforcement. In Proceedings of the 4th International Conference on Big Data Analytics, Data Mining and Computational Intelligence, Porto, Portugal, 16–19 July 2019.
127. Bu, K.; Li, X.; Wang, K.; Li, Y. Data analysis of public food safety cases based on Apriori. In Proceedings of the 2020 Chinese Control and Decision Conference (CCDC), Hefei, China, 22–24 August 2020; pp. 343–348.
128. Tiozzo, B.; Ruzza, M.; Rizzoli, V.; D'Este, L.; Giaretta, M.; Ravarotto, L. Biological, chemical, and nutritional food risks and food safety issues from italian online information sources: web monitoring, content analysis, and data visualization. *J. Med. Internet Res.* **2020**, *22*, e23438. [CrossRef] [PubMed]
129. Liu, J.; Li, Y.; Peng, Y.; Deng, J.; Chen, X. Detection of Food Safety Topics Based on SPLDAs. In Proceedings of the International Conference on Security and Privacy in Communication Networks: 10th International ICST Conference, SecureComm 2014, Beijing, China, 24–26 September 2014; Revised Selected Papers, Part I 10; Springer: Berlin/Heidelberg, Germany, 2015; pp. 551–555.
130. Thakur, M.; Olafsson, S.; Lee, J.S.; Hurburgh, C.R. Data mining for recognizing patterns in foodborne disease outbreaks. *J. Food Eng.* **2010**, *97*, 213–227. [CrossRef]
131. Sadilek, A.; Kautz, H.; DiPrete, L.; Labus, B.; Portman, E.; Teitel, J.; Silenzio, V. Deploying nEmesis: Preventing foodborne illness by data mining social media. *AI Mag.* **2017**, *38*, 37–48. [CrossRef]
132. Harrison, C.; Jorder, M.; Stern, H.; Stavinsky, F.; Reddy, V.; Hanson, H.; Waechter, H.; Lowe, L.; Gravano, L.; Balter, S. Using online reviews by restaurant patrons to identify unreported cases of foodborne illness—New York City, 2012–2013. *Morb. Mortal. Wkly. Rep.* **2014**, *63*, 441.
133. Nsoesie, E.O.; Kluberg, S.A.; Brownstein, J.S. Online reports of foodborne illness capture foods implicated in official foodborne outbreak reports. *Prev. Med.* **2014**, *67*, 264–269. [CrossRef] [PubMed]
134. Effland, T.; Lawson, A.; Balter, S.; Devinney, K.; Reddy, V.; Waechter, H.; Gravano, L.; Hsu, D. Discovering foodborne illness in online restaurant reviews. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1586–1592. [CrossRef] [PubMed]
135. Hu, R.; Zhang, D.; Tao, D.; Hartvigsen, T.; Feng, H.; Rundensteiner, E. TWEET-FID: An Annotated Dataset for Multiple Foodborne Illness Detection Tasks. *arXiv* **2022**, arXiv:2205.10726.
136. Kate, K.; Negi, S.; Kalagnanam, J. Monitoring food safety violation reports from internet forums. In *e-Health–For Continuity of Care*; IOS Press: Amsterdam, The Netherlands, 2014; pp. 1090–1094.
137. Zhang, D.; Zhang, H.; Wei, Z.; Li, Y.; Mao, Z.; He, C.; Ma, H.; Zeng, X.; Xie, X.; Kou, X.; et al. IFoodCloud: A Platform for Real-time Sentiment Analysis of Public Opinion about Food Safety in China. *arXiv* **2021**, arXiv:2102.11033.
138. Qian, C.; Murphy, S.; Orsi, R.; Wiedmann, M. How can AI help improve food safety? *Annu. Rev. Food Sci. Technol.* **2023**, *14*, 517–538. [CrossRef] [PubMed]
139. Ahn, Y.Y.; Ahnert, S.E.; Bagrow, J.P.; Barabási, A.L. Flavor network and the principles of food pairing. *Sci. Rep.* **2011**, *1*, 196. [CrossRef] [PubMed]
140. De Clercq, M.; Stock, M.; De Baets, B.; Waegeman, W. Data-driven recipe completion using machine learning methods. *Trends Food Sci. Technol.* **2016**, *49*, 1–13. [CrossRef]
141. Wiegand, M.; Roth, B.; Klakow, D. Knowledge acquisition with natural language processing in the food domain: Potential and challenges. In Proceedings of the Cooking with Computers Workshop (CwC), Montpellier, France, 28 August 2012; LIRMM: Montpellier, France, 2012; pp. 46–51.
142. Sapienza, S.; Palmirani, M. Emerging data governance issues in big data applications for food safety. In Proceedings of the Electronic Government and the Information Systems Perspective: 7th International Conference, EGOVIS 2018, Regensburg, Germany, 3–5 September 2018; Proceedings 7; Springer: Berlin/Heidelberg, Germany, 2018; pp. 221–230.