Yuzuncu Yil University
Journal of Agricultural Sciences
(Yüzüncü Yıl Üniversitesi Tarım Bilimleri Dergisi)
https://dergipark.org.tr/en/pub/yyutbd

ISSN: 1308-7576
e-ISSN: 1308-7584

Research Article

New Approaches for Food Marketing Strategies and Sectoral Policies: Text Mining and NLP-Based Sentiment Analysis

Yusuf ÇAKMAKÇI*[1], Nuray HELVACI[2], Harun HURMA[3], İdris AKKUZU[4]

[1,2,3]Department of Agricultural Economics, Faculty of Agriculture, Tekirdağ Namık Kemal University, Tekirdağ, Türkiye
[2,4]Department of Political Science and Public Administration, Faculty of Economics and Administrative Sciences, Tekirdağ Namık Kemal University, Tekirdağ, Türkiye

[1]https://orcid.org/0000-0002-5136-9102, [2]https://orcid.org/0009-0003-8602-5099, [3]https://orcid.org/0000-0003-1845-3940
[4]https://orcid.org/0000-0003-1069-9719

*Corresponding author e-mail: cak76yusuf@gmail.com

Article Info

Received: 25.04.2025
Accepted: 24.11.2025
Online published: 15.12.2025
DOI: 10.29133/yyutbd.1683675

Keywords

Consumer Preferences,
Horse Racing Industry,
Natural Language Processing,
Sentiment Analysis,
Text Mining

Abstract: This study investigates the application of Natural Language Processing (NLP) and text mining techniques to enhance the understanding of consumer behavior and sectoral perspectives within the food marketing and horse racing industries. Two primary datasets were utilized: a face-to-face survey conducted with 171 consumers in Tekirdağ, Türkiye, and in-depth interviews with 20 employees from the Turkish horse racing sector. Survey responses were analyzed using R, incorporating text mining methods such as word frequency analysis, bigram identification, chi-square testing, and network analysis. The findings revealed statistically significant associations between purchased foods and specific demographic variables. Notably, household size was significantly associated with cheese consumption ($\chi^2(1) = 6.453$, $p = 0.011$), and gender was significantly related to vegetable consumption ($\chi^2(1) = 4.168$, $p = 0.041$). Additionally, borderline associations were identified between gender and fruit ($p = 0.061$) and egg ($p = 0.080$) consumption, as well as between the number of household workers and yoghurt consumption ($p = 0.054$). Network analysis highlighted the central role of items such as vegetables, fruit, milk, and cheese across various labeling categories, including "organic," "natural," and "cooperative." Interview data were processed in Python using sentiment analysis and clustering techniques. Two primary sentiment-based clusters emerged: one reflecting positive perceptions related to horse care and professional identity, and another indicating dissatisfaction with social life and work-life balance. Overall, the study emphasizes the importance of natural language processing (NLP) and text mining in producing reliable information to influence marketing strategies and policy making in agricultural economics and social science fields.

To Cite: Çakmakçı, Y, Helvacı, N, Hurma, H, Akkuzu İ, 2025. New Approaches for Food Marketing Strategies and Sectoral Policies: Text Mining and NLP-Based Sentiment Analysis. *Yuzuncu Yil University Journal of Agricultural Sciences*, 35(4): 758-775.
DOI: https://doi.org/10.29133/yyutbd.1683675

Footnote: This article utilized data from the master's theses of Çakmakçı (2020) and Helvacı (2024).

1. Introduction

The rapid development of information and communication technologies has led individuals and organisations to generate large amounts of data instantaneously. These masses of data may appear in different formats as structured, semi-structured or unstructured. Since structured and semi-structured data are stored in a certain order, classical analysis methods function effectively in such data. However, unstructured data consists of texts containing valuable information such as user comments, social media posts, interviews with farmers or industry representatives, and market analyses, pushing the limits of traditional methods (Jiang, 2012; Javaid et al., 2024). With the development of artificial intelligence, the processing of data in different quantities and structures has become easier with its intensive use in areas such as education, health, finance and agriculture (Rashid and Kausik, 2024).

At this point, text mining and natural language processing (NLP) methods stand out as powerful tools that provide inferences by analyzing large volumes of text data. With text mining techniques, responses obtained from consumer surveys, user comments, and interviews with industry stakeholders can be systematically evaluated, and new approaches can be developed regarding individuals' decision-making processes and industry dynamics (Soleimanian Gharehchopogh and Abbasi Khalifelu, 2011). Methods such as sentiment analysis, topic modelling and text classification provided by NLP make the strategic decisions of businesses and policy makers more informed by revealing patterns in unstructured data.

Farmer opinions, consumer preferences and sectoral assessments play a critical role in shaping the marketing strategies of businesses and predicting the future potential of sectors (Iftikhar et al., 2021; Ylilehto et al., 2021; Penone et al., 2024). In particular, the food marketing sector is shaped by factors such as changing expectations of consumers, healthy living trends, demand for organic products and sustainability awareness (İnan, 2021; Parashar et al., 2023). On the other hand, the perceptions, job satisfaction, economic expectations and challenges faced by sector employees are of great importance for the sustainability of the sector (Turgut et al., 2012; Dumitru et al., 2023). In addition to traditional market research methods, text mining techniques leverage/utilize textual data obtained from various sources such as social media analyses, consumer comments and survey data (Yüksel et al., 2018). Text mining allows researchers to reveal comprehensive information about consumer preferences using textual data (Alzate et al., 2022). It provides valuable inferences about the strengths and weaknesses of the sector, employee motivations and areas for improvement from the perspective of sector employees (Dinca et al., 2024; AlKhalifa et al., 2025). In the light of this information, it can be said that NLP and text mining have the potential to have a wide range of applications, especially in areas such as social sciences (Hou and Huang, 2025) and agricultural economics. In this context, the study offers an integrated methodological approach that aims to demonstrate the versatile application potential of text mining and Natural Language Processing (NLP) techniques in future research, particularly within the fields of agricultural economics and social sciences, by evaluating data obtained from different sectors within a unified analytical framework.

In the existing literature, there are many studies that demonstrate the effectiveness of text mining methods in analysing consumer behaviour and making sectoral evaluations (Grimmer and Stewart, 2013; Karami et al., 2018; Wang et al., 2023)text mining analyses performed on internet and social media data (Lazer et al., 2014; Topaçan, 2016), determining consumers' perceptions about certain brands or products and optimising marketing strategies (Liu, 2012; Cubukcu Cerası et al., 2024). In the field of agriculture, it is used in food waste management, addressing problems in agricultural production and evaluating agricultural practices (Călin and Coroiu, 2024; Moldovan et al., 2024). On the other hand, there is a gap in NLP and text mining studies that jointly utilize more than one type of data collected by survey and interview methods are used together to address both consumer behavior and sectoral policy issues. This study addresses this gap by utilizing textual data derived from two separate yet complementary sources: consumer survey responses and in-depth interviews with sector employees. These diverse datasets were employed to explore the potential of NLP and text mining methods in analyzing unstructured textual data from both consumers and sector employees, thereby highlighting the comprehensive applicability of these methods in agricultural economics and social sciences (Huang et al., 2022; Delanerolle et al., 2025).

This study investigates the potential of NLP and text mining in the food marketing and horse racing sectors. Crucially, the study's integrated methodological approach, utilizing text data from both structured survey responses and unstructured interviews, is specifically designed to demonstrate the versatile application (versatility) of NLP across distinct data types and disciplinary contexts (Chen, 2023). NLP analysis of interview data and text mining of survey data can provide a better understanding of industry dynamics. These analyses aim to assist businesses in developing marketing strategies and shaping industry policies (OECD, 2023). The study analyses applications in social sciences and agricultural economics, presents case studies and makes recommendations for future research.

2. Natural Language Processing (NLP) and Text Mining: Possibilities for Use in Agricultural Economics and Social Sciences

Natural Language Processing (NLP) and text mining are complementary analysis methods with different focus areas (Kao and Poteet, 2005). NLP is a field of data science developed for computers to understand, interpret and process human language. By focusing on the structural and sentiment features of language, various techniques are used to analyse texts and draw meaningful conclusions. The possibilities offered by NLP play an important role in analysing text data, making inferences and modelling the complexity of language (Basha et al., 2023, Maithili et al., 2023; Stryker and Holdsworth, 2025). Text mining is an analysis method for discovering patterns and relationships in large text data. In this process, meaningful information is extracted from texts using databases, statistical analyses and machine learning techniques (Salloum et al., 2018; De et al., 2022). In this context, NLP stands out as a sub-field used to provide a more in-depth meaning extraction within text mining. In particular, it can be said that NLP focuses more on the structural analysis of language, while text mining enables pattern discovery in large-scale texts.

Survey data, focus group discussions, in-depth interviews, social media posts, sectoral reports and news content are widely used data collection methods in agricultural economics and social sciences research (Nyariki, 2009; Issa, 2015; Sørensen et al., 2022). In this context, qualitative (textual) data can be collected in addition to the quantitative data obtained by the mentioned methods. These textual data allow individuals to express their ideas, opinions and attitudes in their own words without any limitations, which provides the opportunity to obtain more comprehensive data about the situation under consideration. It becomes difficult to evaluate these comprehensive and often complex data with classical analysis methods. In this case, NLP and text mining techniques have the potential to offer various advantages in the evaluation of textual data. Moreover, NLP and text mining techniques facilitate the processing and interpretation of qualitative (textual) data to be collected from researches and provide new perspectives. In particular, textual data obtained from survey data obtained with open-ended questions, focus group discussions (Schnieder et al., 2024), in-depth interviews, social media posts, sectoral reports and news content can be analysed with NLP and text mining techniques to reveal important inferences. These techniques allow extracting meaningful information from texts, identifying keywords and terms, categorising texts thematically and identifying relationships between texts.

NLP techniques, which provide in-depth understanding of texts by modelling language structural and sentiment features, can help identify important policy themes in the context of agricultural economics and social sciences (Atan, 2020). In particular, methods such as sentiment analysis are used to understand public perspectives by examining emotional tones (positive-negative attitudes) in farmer and consumer survey data, social media data, or feedback from product and service users such as farmers (Prasad et al., 2008), consumers (Drury and Almeida, 2011; Surjandari et al., 2015; Whisner et al., 2016; Drury and Roche, 2019; Hegde et al., 2021; Gutiérrez Domínguez et al., 2024; Paleologo et al., 2024). Similarly, text mining techniques provide the possibility of identifying keywords and terms in large datasets, as well as revealing co-occurring words and relationships. This can be particularly useful in analysing textual datasets in agricultural economics and social science reports or farmer and consumer surveys. On the other hand, the widespread use of NLP and text mining applications brings challenges such as language complexity, ambiguities, irony and cultural differences. To overcome these challenges, solutions such as data preprocessing techniques, model selection and multilingual analysis methods can be produced.

In conclusion, NLP and text mining can provide many benefits in agricultural economics and social science research, such as analysing large text data, identifying themes and patterns, providing quick answers to questions and measuring opinions about a topic or product. However, taking into account factors such as language complexity and data quality will allow these methods to be used more effectively in the future for more complex and in-depth analyses.

3. Materials and Methods

3.1. Case 1 study: Consumer purchase of food products marketed under various concepts (face-to-face survey)

This study used primary data collected through face-to-face surveys with 171 consumers in Süleymanpaşa, Tekirdağ in 2019-2020 within the scope of Çakmakçı's master's thesis (Çakmakçı, 2020). The minimum required sample size (n=171) was calculated using the proportional sampling formula for a known population, based on a 95% confidence level and a 7.5% margin of error. This calculation approach, which assumes maximum variance (p=0.50, q=0.50), adheres to standard statistical practices in social science research (Yüzbaşıoğlu and Kaplan, 2019; Krejcie & Morgan, 1970). In addition to the demographic-economic variables of consumers, the study handled the data set consisting of food names obtained from open-ended questions asking consumers to specify the foods they bought under the categories of 'natural', 'organic', 'good agricultural practices', 'direct from the producer' and 'co-operative branded' food (Table 1). All analyses were performed using the R programming language (version 4.4.1) and related packages. R was chosen because of its statistical analysis capabilities, especially categorical data association tests, and the integrated workflow offered by the tidyverse ecosystem in data processing and text mining steps (Wickham et al., 2019; R Core Team, 2024).

Table 1. An example of consumer data

No	Age	Gender	Marriage	Family size	Worker	Education	Income	Food expenses	Additional Income	Child
1	55	male	married	3	2	Pre and Undergraduate	3500	2000	no	no
2	27	female	Single	3	2	Pre-and Undergraduate	2500	1000	yes	yes
3	65	male	married	3	3	Post graduate	5700	650	no	no

Know Environment Product	Home Status	Natural	Organic	Gap	Producer	Sentence Kooperative
yes	rented	Olive Oil	Fruit, Vegetable	Olive, Chicken, Tomato	Cheese, Olive	honey, Hazelnut, Pasta
yes	rented	Cheese	Egg, honey	olive		
yes	rented				Yogurt, Bread, Egg	Grape molasses, Olive oil

During the analysis process, the survey data were read (readxl), each respondent was assigned an ID, textual responses were translated into English, merged and subjected to extensive pre-processing. The data were converted into long format and food items were tokenised. During this pre-processing stage, missing values (NaN) within the textual data columns were systematically identified and removed to ensure data integrity for subsequent frequency and network analyses. Texts were cleaned (stringr) and, as a critical step for consistency of analysis, food items with different spellings were manually converted into a standard terminology. Prior to word-based analyses, English stopwords were extracted using the 'tm' package (Feinerer et al., 2008). Frequencies of standardised food names were calculated and visualised using ggplot2 and wordcloud packages (Wickham, 2016; Fellows, 2018). In addition, as an exploratory analysis on the original texts, consecutive word pairs (bigrams) were identified using the tidytext package (Silge and Robinson, 2016). For demographic analyses, data consisting of standardised food names were combined with demographic information. Continuous demographic variables were divided into categorical groups and food preferences of different subgroups were visualised comparatively using the ggplot2 package. In compliance with the rigorous standards required for publication, the categorical variables utilized in the comparative visualisations and tabular presentations (e.g., 'Education' and 'Income' subgroups) were systematically reviewed. This process ensures the enhanced clarity and accurate terminological consistency of the visual evidence presented in the findings section, thereby eliminating potential ambiguities in interpretation by the reader. Finally, the relationship patterns between food items co-mentioned by the same respondent were examined by network analysis; co-mention frequencies were calculated with the widyr package, an undirected network graph was created with the igraph package and visualised with the ggraph package (Csardi and Nepusz, 2006; Robinson and Silge, 2022; Pedersen, 2023). Network visualisation was performed using the Kamada-Kawai layout algorithm (Kamada and Kawai, 1989), which aesthetically reveals the connections and clusters between nodes. This arrangement aims to position the nodes in the network at optimum distances, while node size represents frequency/centrality, colour represents category and edge thickness realistically represents strength of association. The tidyverse collection, readxl, tidytext, tm, wordcloud, igraph, ggraph and widyr packages were used in the analyses (R Core Team, 2024).

3.2. Case 2 study: Evaluation of the sector by horse racing sector employees in Türkiye (In-depth interview)

This study analysed a qualitative dataset from Helvacı's master's thesis, reflecting the experiences of racehorse industry workers in Türkiye (Helvacı, 2024). The dataset consisted of manual transcriptions of in-depth interviews with 20 industry insiders, which were combined into a single chunk of text for analyses (Table 2). The size of this qualitative sample was specifically justified by the principle of data saturation (Glaser and Strauss, 1967). Data collection was stopped after the 20th interview, as no new themes or significant information emerged, thereby confirming saturation and providing sufficient depth and reliability for the analysis (Guest et al., 2006; Malterud et al., 2016). All analyses were performed in the Python Jupyter environment. Python was chosen for this study because it allows us to efficiently manage multi-step analysis tasks such as text vectorisation (TF-IDF), natural language processing (TextBlob, sentiment analysis with VADER), dimensionality reduction (PCA) and clustering (KMeans) with a comprehensive library ecosystem that offers integrated tools such as scikit-learn.

Table 2. Example excerpts from participants' interview transcripts

Participant ID	Combined Interview Responses
1	I used to work as a stable hand for a year. I worked in a large stable with 12 horses. I would make one horse's bed, give water to another. The master would say "run," and I would run. Actually, you're doing an internship? First, you'll go and work as a stable hand. You'll learn the horses' temperaments, gain confidence, and then observe the horse's condition, movements, and behavior.
2	………………

For text mining and sentiment analysis, the data were preprocessed through a preprocessing process including lower case conversion, tokenisation, stop word cleaning, non-letter character extraction and stemming (lemmatisation) (Vijayarani and Janani, 2016). The preprocessed text was vectorised using the TF-IDF method for word representation (Salton and Buckley, 1988). Frequently used terms were visualised with a word cloud (Viégas and Wattenberg, 2008). Sentiment polarity at sentence level was calculated using the TextBlob library (Loria, 2018). Principal Component Analysis (PCA) was applied to the TF-IDF vectors for dimensionality reduction (Jolliffe and Cadima, 2016). The obtained two-dimensional data were clustered with KMeans algorithm (Hartigan and Wong, 1979), and the optimal number of clusters was determined using Elbow Method and Silhouette Score (Rousseeuw, 1987). The overall emotional tone of each cluster was labelled with VADER (Hutto and Gilbert, 2014). Finally, the results of the sentiment analysis by clustering were visualised through a scatter plot on the PCA-reduced data, with coloured emojis representing the sentiment labels for each cluster.

4. Results and Discussion

4.1. Case 1: Consumer purchase of food products marketed under various concepts (face-to-face survey) results

Figure 1 shows the most frequently purchased food products (A) and the word cloud (B). It is seen that the top 10 most purchased foods in all food categories by consumers are 'egg', 'milk', 'fruit', 'cheese', 'vegetable', 'chicken', 'honey', 'yogurt', 'rice', 'oil' (Figure 6.1-A). Similarly, Figure 6.1-B is the wordcloud image of the first 200 foods that are said to be purchased most by consumers. In the light of these distributions, frequency distribution and wordcloud visualisations can be considered as an important visual tool in the analysis of consumption trends and food preferences.

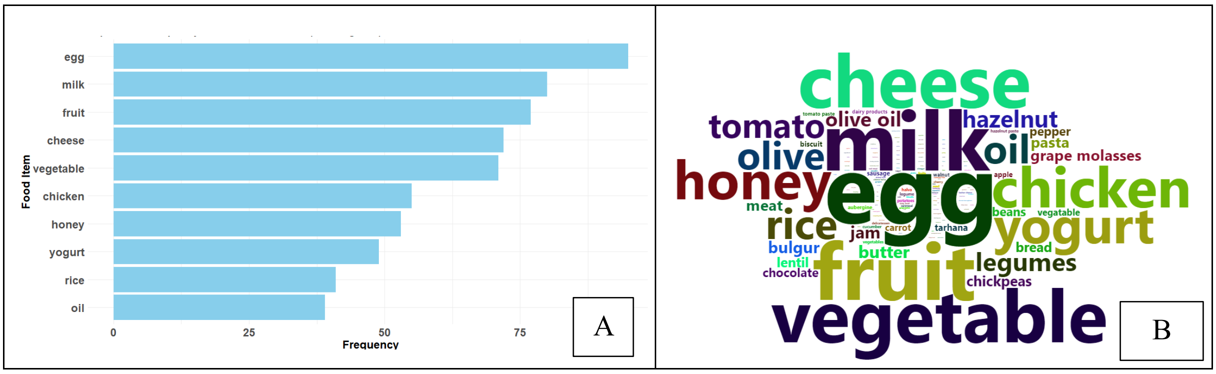

Figure 1. Top 10 most frequently mentioned foods purchased in all food categories (A); word cloud of 200 food products (B).

The graphs in Figure 2 present the most frequent word pairs (bigrams) and their frequencies for the different food categories ('Natural', 'Organic', 'Good Agricultural Practices - GAP', 'Direct from Producer', 'Cooperative branded foods') and for the texts where all data is combined ('All Foods'); bigram analysis aims to reveal the product pairs or descriptive phrases most frequently associated with each category. According to the results of the analysis, dairy product combinations such as 'milk yoghurt', 'yoghurt cheese', 'egg milk' are prominent in the Natural Foods category. The most frequently associated pairs with Organic Foods were 'egg and milk', 'olive oil', 'grape molasses' and 'fruit and vegetable'. While 'vegetables and fruits' is the most dominant pair in the GAP Foods category, specific product combinations such as 'fruit and chicken' and 'chicken and chilli' are also noteworthy. Under the heading of Foods Purchased Directly from Producers, 'fruit and vegetables' and 'fruit and vegetables' are by far the most frequently mentioned pairs, emphasising the emphasis on fresh products, while there are also pairs such as 'fruit and cheese' and 'egg and cheese'. Co-operative Foods is most prominently characterised by 'olive oil' and its derivatives ('olive oil'), while 'fat cheese', 'rice oil' and pairs containing nuts ('hazelnut paste', 'hazelnut rice') are also associated with this category. An overview of the All Foods (Combined) data shows that 'fruit and vegetables', 'olive oil' and 'egg and milk' are the main product associations that are frequently encountered in different categories and are generally important. The findings indicate that consumers frequently associate milk and dairy products within the Natural and Organic product categories, reflecting a broader tendency to consume items such as milk, yoghurt, and cheese in combination an observation also supported by regional studies, such as the one conducted in Erzincan (Çebi et al., 2018). Furthermore, the prominent co-occurrence of fresh produce and dairy products in the "Direct from Producer" and "Cooperative" categories underscores the continued significance of traditional supply channels as a key determinant in consumer preference for these food types.

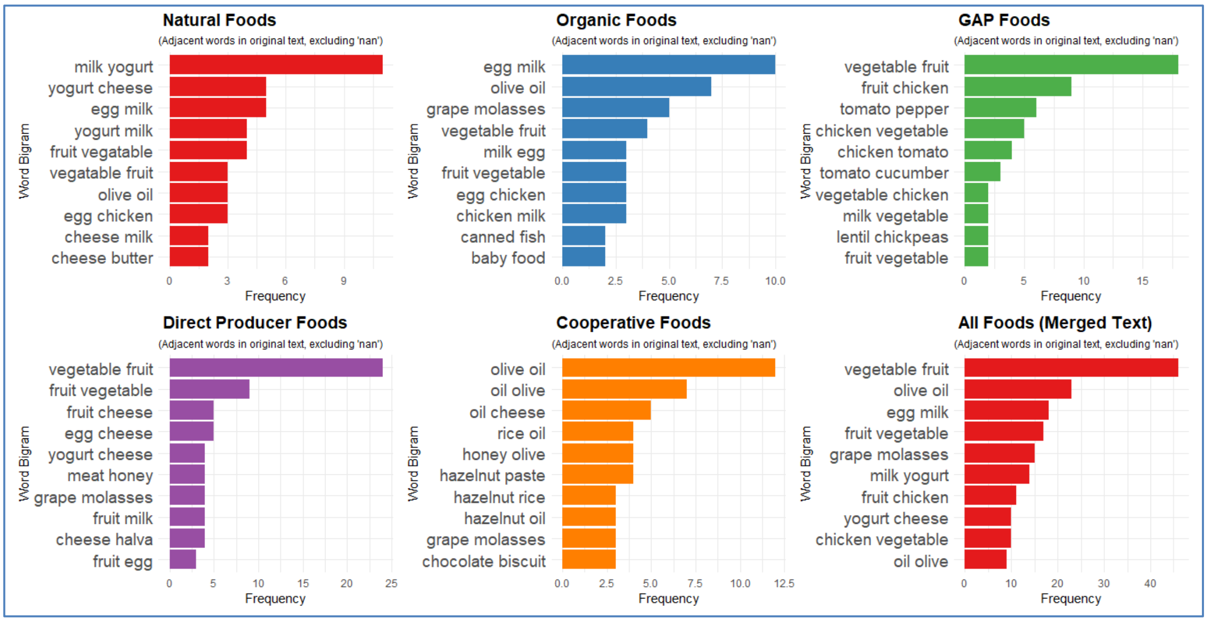

Figure 2. Frequency distribution of products purchased in all food categories and aggregated data (Top 10 bigrams).

Figure 3 visualises the proportional differences in consumers' food preferences according to various socio-demographic factors (age, gender, etc.). These graphs indicate trends, for example, that there are noticeable differences in the consumption rates of certain food products between age groups. However, Chi-square analyses (Table 3) were conducted to determine whether these observed differences are statistically significant. As a result of these analyses, statistically significant relationships were found only between household size and cheese consumption ($\chi^2(1) = 6.453$, $p = 0.011$) and between gender and vegetable consumption ($\chi^2(1) = 4.168$, $p = 0.041$). In addition, the relationships between the number of employees and yoghurt consumption ($p = 0.054$), gender and fruit consumption ($p = 0.061$), gender and egg consumption ($p = 0.080$), and Environmental product knowledge and milk ($p = 0.094$) were found to be borderline significant at the 10% significance level. The findings of a study conducted in Iğdır province indicated a correlation between the purchase of organic food products and socioeconomic variables, including "household size," "monthly household income," and "educational attainment," which influence organic food purchasing behavior (Kadirhanoğulları et al., 2021). Furthermore, findings from another study suggest that socioeconomic variables characterizing consumer profiles have statistically significant influences on food preferences (İnan, 2021; Kaçmaz et al., 2023). As a result, such detailed analyses of text mining can provide valuable data for developing marketing strategies and policies for the agriculture and food sector. As an important methodological implication, these findings reveal that relying solely on frequency or ratio graphs can be misleading, and confirming the observed trends with statistical tests is critical to reach more reliable conclusions.

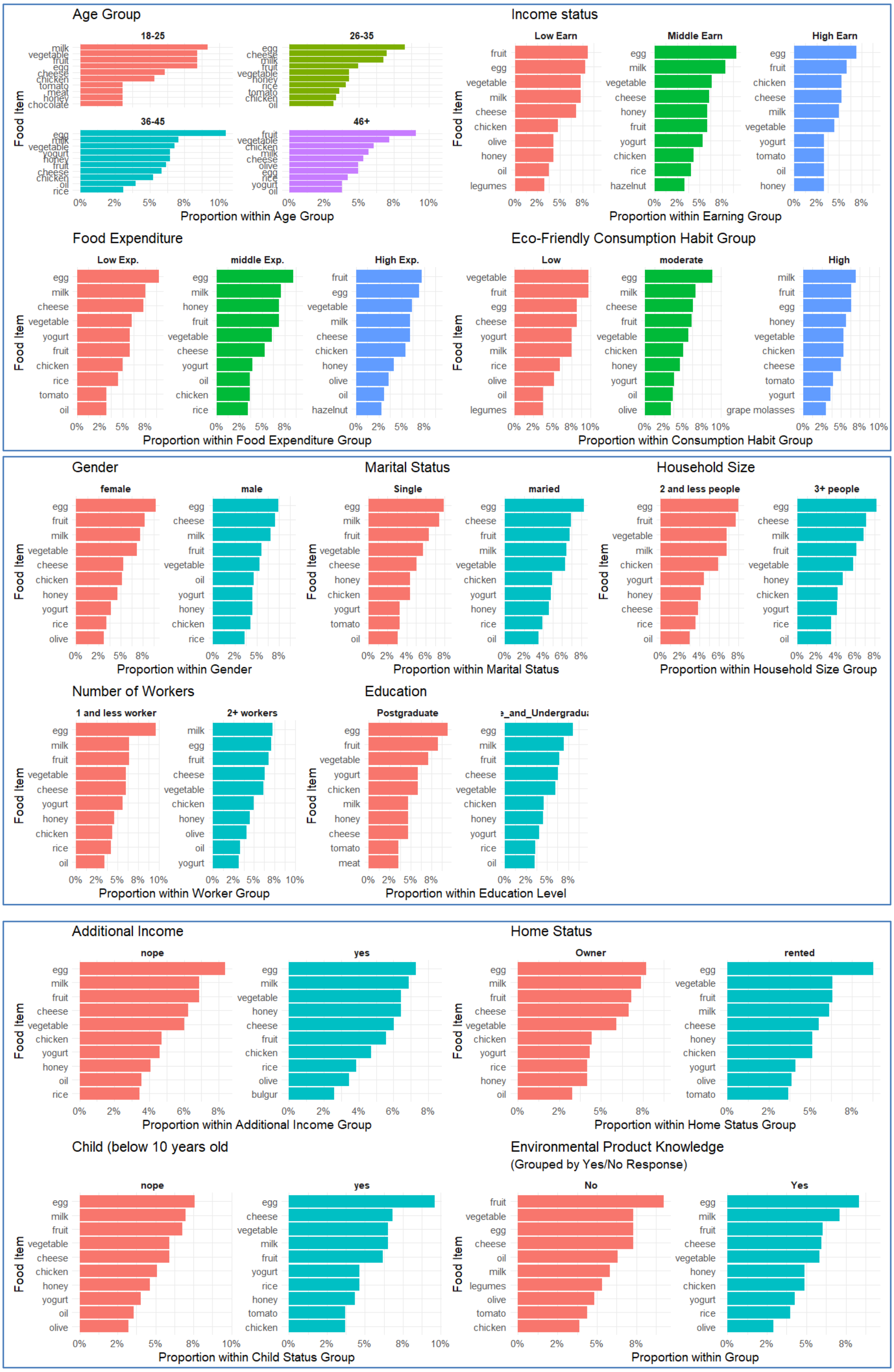

Figure 3. Comparison of purchased natural food labelled products with consumer demographic data.

Table 3. Chi-Square test results of the relationships between demographic variables and food product consumption

Demographic Variable	Purchased Foods	N	Df	Test Statistic (Chi-Sq)	P Value
Household size	cheese	171	1	6.453	0.011**
Gender	vegetable	171	1	4.168	0.041**
Number of workers	yogurt	171	1	3.717	0.054*
Gender	fruit	171	1	3.512	0.061*
Gender	egg	171	1	3.060	0.080*
Environmental product knowledge	milk	171	1	2.805	0.094*

Note: ($p < 0.01$***, $p < 0.05$**, $p < 0.10$*).

Figure 4 uses network analysis, visualized with the Kamada-Kawai algorithm, to explore relationships between consumer food preferences across categories. In this network, foods are nodes, their associations are edges, and node size reflects popularity. Staple foods like 'vegetable', 'fruit', 'cheese', and 'milk' demonstrate high centrality. Different categories are color-coded: blue for organic products, pink for product direct-from-producer, orange for cooperative products, red for natural products, and green for GAP products (good agricultural practices). Organic foods (blue) strongly link with vegetables, fruits, dairy, and cereals, reflecting health and environmental awareness influences mentioned in the literature. Direct-from-producer foods (pink) connect more with fermented products like cheese, honey, and yoghurt, perceived by consumers as traditional and natural. Co-operative branded foods (orange) associate with staples like legumes, tomatoes, and honey, showing a wide network reach potentially linked to trust and positive health perceptions (Çakmakçı, 2020; Yüzbaşıoğlu, 2021). The large size and centrality of 'vegetable' and 'fruit' nodes highlight their frequent preference and function as bridges connecting various labelling formats (organic, natural, direct) (Taysi et al., 2021; Parashar et al., 2023; Acibuca and Kaya, 2024). These findings underscore the utility of network analysis in understanding consumer trends and food category relationships, providing a basis for future research into consumer segmentation and sustainable food systems.

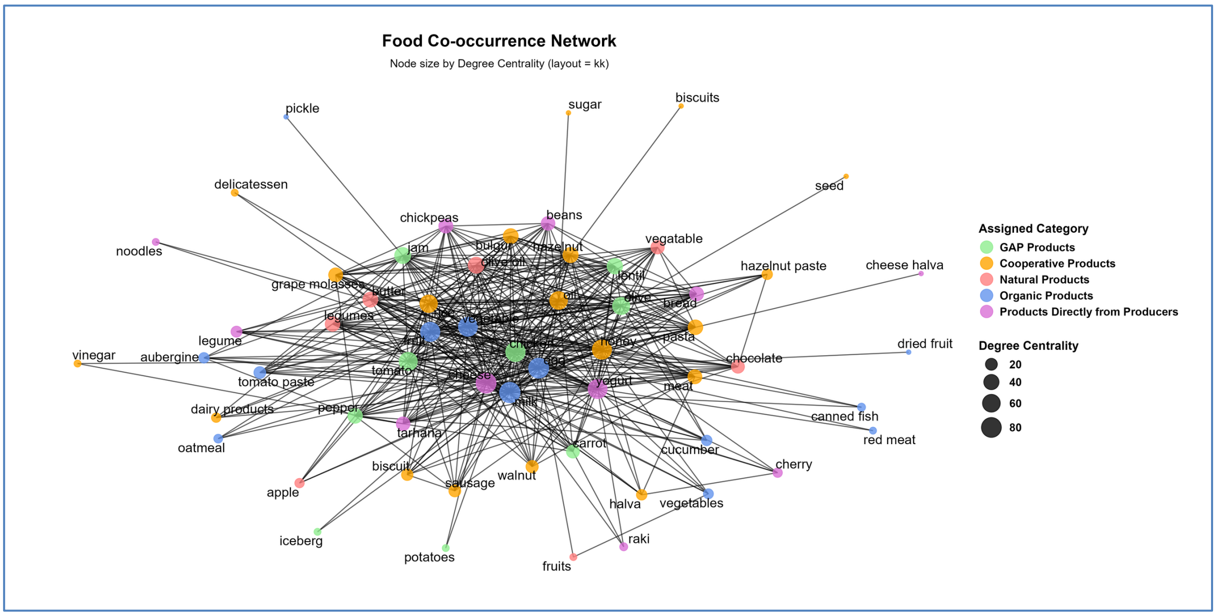

Figure 4. Co-occurrence of purchased foods based on consumer responses Purchase Network Edge Weights (layout=kk).

4.2. Case 2: Evaluation of the sector by horse racing sector employees in Türkiye (In-depth interview) results

The total number of sentences obtained from the interviews conducted in the horse racing industry was found to be 255 with the help of TextBlob, a popular natural language processing (NLP) library developed for the Python programming language. It was observed that 65 of these sentences were positive, 40 were negative and 150 were neutral. The polarity of the most positive sentence was calculated as (0.80) and the polarity of the most negative sentence as (-1.00) (table 4).

Table 4. Basic statistics and TextBlob-based polarity distribution of interview data

Sentiment Label	Average polarity	Minimum polarity	Maximum polarity	Number of Sentences
Positive	0.26	0.05	0.80	65
Negative	-0.26	-1.00	-0.06	40
Neutral	0.00	-0.03	0.05	150
Total				255

The word cloud presented in Figure 5 was created by analysing the textual data obtained from the evaluations of horse racing sector employees in Türkiye by using text mining techniques. The word cloud visualises the most frequent words in the text and their relative importance. Prominent terms include ‘horse,’ ‘groom,’ Concepts related to the equine industry and employment, such as ‘work,’ ‘job,’ ‘jockey,’ ‘training,’ and ‘care,’ are included. Thus, the word cloud is effective to reveal the main themes of the textual data and to visualize the topic distribution.

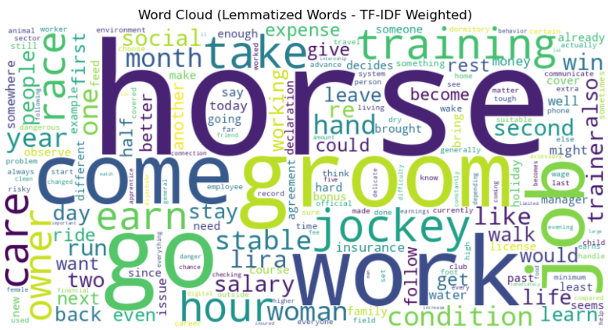

Figure 5. The most frequently used words throughout the study.

Figure 6 shows the results of the Elbow Method and Silhouette Score for determining the most appropriate number of clusters according to the sentiment scores obtained from the evaluations of employees in the horse racing sector in Türkiye. In both methods, a significant break and the highest Silhouette Score (0.509) was observed in the k = 2 cluster, which reveals that grouping employee evaluations in 2 main clusters is the most appropriate approach. Since the Elbow Method evaluates intra-cluster variance, while the Silhouette Score measures cluster separation, their combined use provides a more reliable determination of the number of clusters.

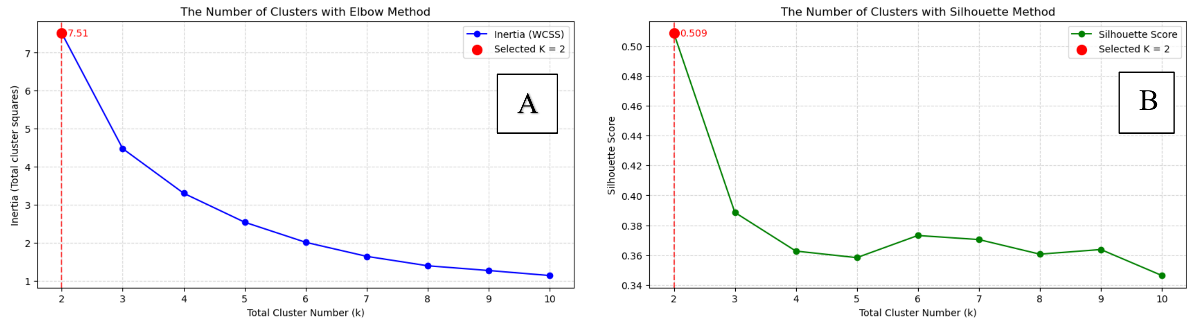

Figure 6. Study ideal cluster number determination A: Elbow method; B: Silhouette method.

Figure 7 analyses the work experiences, perceptions, and emotional states of individuals working in the racehorse breeding sector through text mining (Atan, 2020), which is frequently used to extract in-depth meaning from qualitative data (Salloum et al., 2018; Schnieder et al., 2024), and sentiment analysis (Liu, 2012; Hutto and Gilbert, 2014), which is a powerful method for measuring individuals' attitudes. Sentiment analysis, especially on text data obtained through direct data collection methods such as interviews (Nyariki, 2009; Issa, 2015), allows for a quantitative assessment of the delicate balance between professional satisfaction and challenges faced by employees. When the figure shows the sentiment scores of each sentence, the calculated mean polarity score (mean polarity score) of 0.03 reveals that, in general, sector employees have a neutral or slightly positive tendency in all of their speech. The statement with the highest positive sentiment score, “When you win, you don’t even think about the money at that moment; you think about the excitement,” emphasises the importance of intrinsic motivation and satisfaction provided by the job by showing that when a race is won, the focus is on the excitement of the moment rather than the financial reward. On the other hand, the statement with the most negative emotion score, “We wake up at 4:00, the worst time to be awake,” indicates that waking up very early in the morning is one of the most challenging aspects of the job. This draws attention to the physically taxing working conditions that are common in agriculture and livestock-related sectors (Dinca et al., 2024). These fluctuations in the overall distribution of emotions indicate that the sector employees expressed both the sources of occupational satisfaction and the physical and mental difficulties they face together in their statements, that is, their experiences are multidimensional. In conclusion, sentiment analysis reveals that it is a valuable method (Liu, 2012; Topaçan, 2016) to reveal the sectoral perspectives of the workforce in private sectors such as racehorse breeding through their own expressions. Moreover, such detailed analyses can provide an important source of data to develop evidence-based policies to improve employee well-being and organisational commitment (Turgut et al., 2012), especially in issues such as regulating working hours and improving employees’ work-life balance, taking into account sectoral conditions in Türkiye (Helvacı, 2024). Such improvements can play a critical role in increasing sustainability and employee satisfaction in the sector.

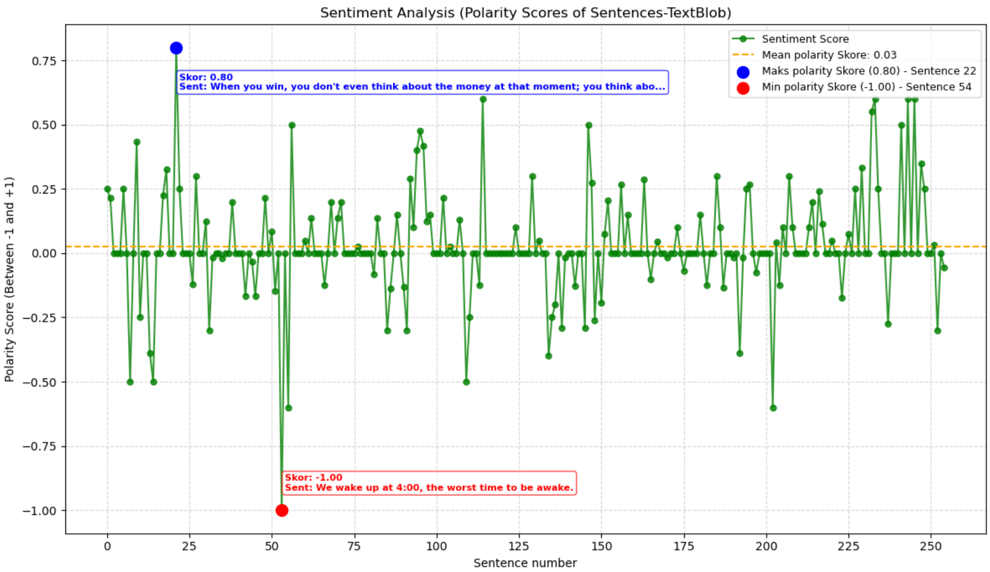

Figure 7. Sentiment analysis status of sentences (TextBlob polarity score).

Figure 8 reveals the trends regarding professional satisfaction, economic gain, and social life balance among individuals working in the racehorse breeding sector using qualitative data mining (Atan, 2020) and sentiment analysis (Liu, 2012; Hutto and Gilbert, 2014) techniques, which are increasingly used to understand stakeholder perspectives (Drury and Roche, 2019; Gutiérrez Domínguez et al., 2024). Applied to identify latent themes in the text data (Vijayarani and Janani, 2016), the K-Means clustering algorithm (Hartigan and Wong, 1979) yielded two main clusters: "Horse/Work/Groom" (Cluster 1) and "Social/Life/Job" (Cluster 2). When examining the sentiment states of the clusters, it was observed that in the first cluster, employees expressed positive (positive) views (polarity score=1) about the topics of horses, work, and grooming; this indicates that employees have a positive attitude towards the fundamental elements of their profession. On the other hand, it is understood that the second cluster concentrates on topics such as social life, life in general, and leaving the job, and employees are markedly dissatisfied (negative) with these issues (polarity score=-99). This points to serious problems in work-life balance and social support, which are considered critical for employee well-being and commitment in the literature (Al Khalifa et al., 2025). This cluster finding necessitates a deeper discussion within the framework of the Effort-Reward Imbalance (ERI) Model. The ERI model posits that chronic stress and dissatisfaction arise when the high effort expended by employees (e.g., long working hours, continuous responsibility for live animals) is not adequately matched by rewards (e.g., fair compensation, appreciation, job security) (Siegrist, 2012). The frequent appearance of terms such as 'hours,' 'load/burden,' and 'family' in our cluster analysis empirically validates that workers in this sector experience a systematic ERI. Furthermore, these findings are complementarily illuminated by the Job Demands-Resources (JD-R) Model. JD-R theory suggests that the combination of high job demands (e.g., long hours, intense workload) and a lack of job resources (e.g., low autonomy, poor social support, insufficient training) negatively impacts employee well-being (Bakker and Demerouti, 2017; Hsu et al., 2025). This situation shows that sector employees have a satisfying experience regarding their profession but are dissatisfied with work-life balance and social life. These findings revealed that employees in the Turkish racehorse sector have a positive attitude towards their profession, while their dissatisfaction with social life and work-life balance was demonstrated through sentiment analysis and clustering techniques. It highlights the critical importance of developing improvement strategies focused on social life and work-life balance, taking into account the specific conditions of the horse racing industry in Türkiye (Helvacı, 2024), in order to increase employees' overall well-being and organizational commitment (Turgut et al., 2012). The findings of the study reveal that NLP and text mining methods have the potential to guide decision-making mechanisms on the mentioned issues.

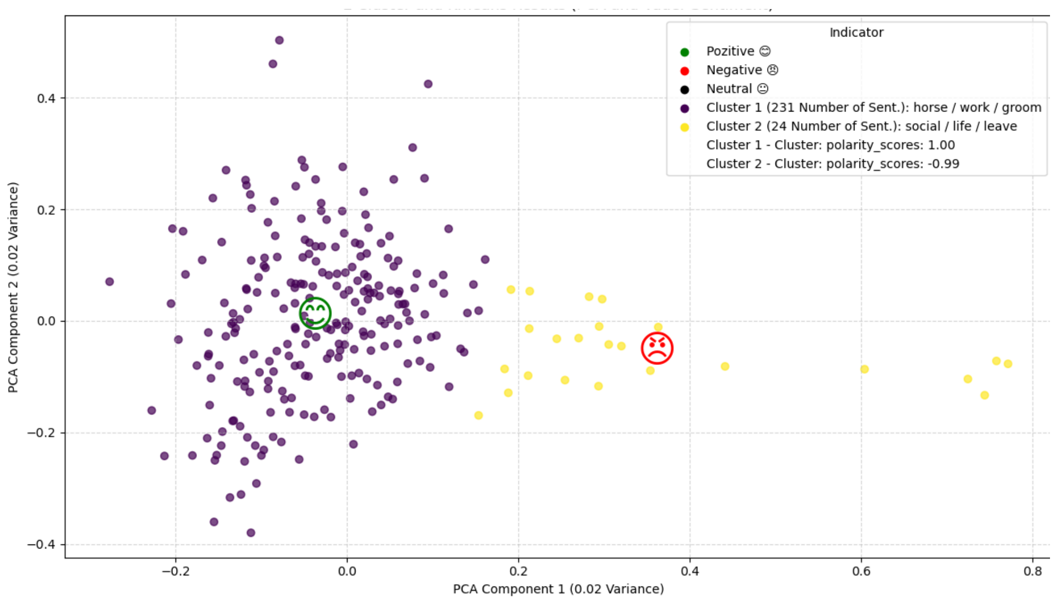

Figure 8. K-means clustering results (with Cluster sentiment scores).

Conclusion

This study demonstrated that the analysis of textual data obtained from consumer surveys through text mining can provide valuable insights into food consumption habits. Using word clouds, frequency distributions (bigrams), and network analyses, consumer purchasing behaviors and preferences associated with demographic characteristics were revealed. Network analysis was found to be particularly effective in identifying the co-purchase tendencies of food products. Although the size of the survey sample and the focus on word-level analysis present limitations, the use of sentence-level data and larger samples in future studies could enhance the generalizability of the findings. Text mining has the potential to contribute to areas such as improving marketing strategies, product development, and promoting healthy eating. Furthermore, it forms a basis for creating consumer segmentations and developing sustainable food consumption strategies.

On the other hand, sentiment analysis, one of the Natural Language Processing (NLP) methods, was used to analyze the perceptions of employees in the horse racing sector. This demonstrated the potential of textual data collected in sectoral evaluations to extract valuable information with the help of NLP. The findings revealed that while employees' professional satisfaction is high, there is a need for improvements in work-life balance and social life. This indicates that sentiment analysis can be a valuable tool in determining sector policies, enhancing employee well-being, and developing sustainable business models. The study results show that NLP-based analyses can contribute to the evaluation of sectoral structure and the development of policies. These results serve as an important tool to guide sector evaluators and food policy makers. Furthermore, future studies could combine sentiment analysis with more complex algorithms such as machine learning and deep learning to derive deeper insights into the attitudes and behaviors of consumers and sector employees.

In this regard, the joint analysis of data from different sectors illustrates the potential of NLP and text mining not only for addressing economic concerns such as consumer behavior and sustainable food systems, but also for exploring critical social dimensions such as job satisfaction and professional identity. These results underscore the value of a unified methodological approach and support the use of text-based analytics in both agricultural economics and social science research for policy development, sectoral evaluation, and strategic planning.

Ethical Statement

Ethical approval for this study was obtained from the Scientific Research and Publication Ethics Committee of the Faculty of Social and Human Sciences, Tekirdağ Namık Kemal University (Approval Number: 455254, Date: 11.06.2024).

The section of this article derived from Yusuf Çakmakçı's thesis is based on data collected before 2020. In accordance with the TR Dizin Ethical Principles (2020), retrospective ethics approval is not required for studies using pre-2020 thesis data. This research does not involve any clinical or experimental procedures on humans or animals; therefore, ethics committee approval was not required.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Funding Statement

No external financing support was received for this research.

Author Contributions

The concept and design of the study: YÇ, HH; sample collection: YÇ, HH, NH; İA data analysis and interpretation: YÇ, HH; statistical analysis: YÇ, HH; visualization: YÇ, HH; manuscript writing: YÇ, NH, HH, İA; All authors read and review final draft.

Acknowledgements

The authors sincerely thank all participants who voluntarily took part in the questionnaire survey and in-depth interviews for their valuable time and contributions to this study.

References

Acibuca, V., & Kaya, A. (2024). Environmental awareness and attitudes of university students: The case of Türkiye. *ISPEC Journal of Agricultural Sciences, 8*(1), 237–246. https://doi.org/10.5281/zenodo.10852912.

AlKhalifa, J., Nawaz, N., & Sawaya, R. (2025). *Motivational factors and employee performance: A comprehensive analysis.* In A. Hamdan (Ed.), Achieving sustainable business through AI, technology education and computer science. *Studies in Big Data*, 158, Springer, Cham. https://doi.org/10.1007/978-3-031-70855-8_53.

Alzate, M., Arce-Urriza, M., ve Cebollada, J. (2022*). Mining the text of online consumer reviews to analyze brand image and brand positioning. Journal of Retailing and Consumer Services***,** *67*, 102989. https://doi.org/10.1016/j.jretconser.2022.102989.

Atan, S. (2020). Metin Madenciliği: İmkânlar, Yöntemler Ve Kisitlar. Mehmet Akif Ersoy University *Journal of Social Sciences Institute, 31*, 220-239. https://doi.org/10.20875/makusobed.476524.

Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, *22*(3), 273–285. https://doi.org/10.1037/ocp0000056.

Basha, M. J., Vijayakumar, S., Jayashankari, J., Alawadi, A. H., & Durdona, P. (2023). *Advancements in natural language processing for text understanding*. In ICONECT-2023, E3S Web of Conferences, *399*, 04031. https://doi.org/10.1051/e3sconf/202339904031.

Çakmakçı, Y. (2020). Determination of consumer perception on eco-friendly concepts used in the marketing of various food products (The case of Tekirdağ/Süleymanpaşa). (Master's thesis), Tekirdağ Namık Kemal University, Institute of Social Sciences, Tekirdağ, Türkiye.

Călin, A. D., & Coroiu, A. M. (2024). What is the attitude of romanian smallholders towards a ground mole infestation? A study using topic modelling and sentiment analysis on social media and blog discussions. *Animals*, *14*(24), 3611. https://doi.org/10.3390/ani14243611.

Çebi, K., Özyürek, S., & Türkyılmaz, D. (2018). The factors affecting consumer choices in dairy products: the case of Erzincan. *Yuzuncu Yıl University Journal of Agricultural Sciences*, *28*(1), 70-77. https://doi.org/10.29133/yyutbd.348206.

Chen, C. W. (2023). A Feasibility Discussion: Is ML suitable for predicting sustainable patterns in consumer product preferences. *Sustainability*, *15*(5), 3983. https://doi.org/10.3390/su15053983.

Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research.

Cubukcu Cerası, C., Balcıoğlu, Y. S., Huseynov, F., & Kılıç, A. (2024). A comprehensive text mining application to understand social media's impact on consumer perception of green consumption. *Bilgisayar Bilimleri ve Mühendisliği Dergisi,* 28-37.

De, S., Dey, S., Bhatia, S., & Bhattacharyya, S. (2022). *Chapter 1 - An introduction to data mining in social networks*. In S. De, S. Dey, S. Bhattacharyya, & S. Bhatia (Eds.), Advanced data mining tools and methods for social computing (pp. 1–25). Academic Press. https://doi.org/10.1016/B978-0-32-385708-6.00008-4

Delanerolle, G., Bouchareb, Y., Shetty, S., Cavalini, H., & Phiri, P. (2025). A pilot study using natural language processing to explore textual electronic mental healthcare data. *Informatics*, *12*(1), 28. https://doi.org/10.3390/informatics12010028.

Dinca, V. M., Trocinescu, B., Stamule, S., Bunea, M. & Dinu, V. (2024). Opportunities and challenges for managers within the East-European agriculture sector: Case study on Romania. *E&M Economics and Management*, *27*(4), 121–134. https://doi.org/10.15240/tul/001/2024-4-008.

Drury, B., Almeida, J. J. (2011). *Identification of fine grained feature based event and sentiment phrases from business news stories*. In Proceedings of the Association for Computing Machinery. Association for Computing Machinery. https://doi.org/10.1145/1988688.1988720.

Drury, B., & Roche, M. (2019). A survey of the applications of text mining for agriculture. *Computers and Electronics in Agriculture, 163*, 104864. https://doi.org/10.1016/j.compag.2019.104864.

Dumitru, E. A., Sterie, C. M., Rodino, S., & Butu, M. (2023). Consumer preferences in the purchase of agri-food products: Implications for the development of family farms. *Agriculture*, *13*(8), 1478. https://doi.org/10.3390/agriculture13081478.

Feinerer, I., Hornik, K., & Meyer, D. (2008). *Text mining infrastructure in R*. Journal of Statistical Software, *25*(5), 1–54. https://doi.org/10.18637/jss.v025.i05.

Fellows, I. (2018). *wordcloud: Functionality to create pretty word clouds, visualize differences and similarity between documents, and avoid over-plotting in scatter plots with text* (R package version 2.6). CRAN. http://blog.fellstat.com/?cat=11.

Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory: Strategies for qualitative research. Aldine Publishing Company. http://www.sxf.uevora.pt/wp-content/uploads/2013/03/Glaser_1967.pdf.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. https://doi.org/10.1093/pan/mps028.

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough?: An experiment with data saturation and variability. *Field Methods, 18*(1), 59–82. https://doi.org/10.1177/1525822X05279903.

Gutiérrez Domínguez, A., Roig-Tierno, N., Chaparro-Banegas, N., & García-Álvarez-Coque, J.-M. (2024). Natural language processing of social network data for the evaluation of agricultural and rural policies. *Journal of Rural Studies*, *109*, 103341. https://doi.org/10.1016/j.jrurstud.2024.103341.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: a k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 28*(1), 100-108.

Hegde, G., Hulipalled, V. R., & Simha, J. B. (2021). *Price prediction of agriculture commodities using machine learning and NLP*. In 2021 Second International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE) 1–6. IEEE. https://doi.org/10.1109/ICSTCEE54422.2021.9708582.

Helvacı, N. (2024). Socio-economic analysis of the horse racing industry in Turkey (Master’s thesis). Tekirdağ Namık Kemal University, Institute of Social Sciences, Tekirdağ, Türkiye.

Hou, Y., & Huang, J. (2025). Natural language processing for social science research: A comprehensive review. *Chinese Journal of Sociology*, *11*(1), 121–157. https://doi.org/10.1177/2057150X241306780.

Hsu, S., Gligor, D., Garg, V., Gölgeci, I., & Choi, R. J. (2025). Exploring supply chain capabilities as drivers for willingness to adopt blockchain technology using a technology–organization–environment (TOE) framework. *Production Planning & Control, 36*(16), 2382–2398. https://doi.org/10.1080/09537287.2025.2533186.

Huang, Y., Wang, X., Wang, R., & Min, J. (2022). Analysis and recognition of food safety problems in online ordering based on reviews text mining. *Wireless Communications and Mobile Computing*, *2022*, 4209732. https://doi.org/10.1155/2022/4209732.

Hutto, C. J., & Gilbert, E. E. (2014). Vader: a parsimonious rule-based model for sentiment analysis of social media text [Paper presentation]. Eighth International AAAI Conference on Weblogs and Social Media (ICWSM-14).

Iftikhar, A., Purvis, L., & Giannoccaro, I. (2021). A meta-analytical review of antecedents and outcomes of firm resilience. *Journal of Business Research*, *135*, 408–425. https://doi.org/10.1016/j.jbusres.2021.06.048.

İnan, R. Assessment of Consumers of Organic Food Purchase Behavior and Attitudes. *Journal of Tourism and Gastronomy Studies*, *9*(1), 220-235. https://doi.org/10.21325/JOTAGS.2021.786

ISSA, F. (2015). Methods of Data Collection in Agricultural Extension Research. AESON.

Javaid, M., Haleem, A., Singh, R. P., & Sinha, A. K. (2024). Digital economy to improve the culture of industry 4.0: A study on features, implementation and challenges. *Green Technologies and Sustainability*, *2*(2), 100083. https://doi.org/10.1016/j.grets.2024.100083

Jiang, J. (2012). Information Extraction from Text. In: Aggarwal, C., Zhai, C. (eds) Mining Text Data. *Springer*, Boston, MA. https://doi.org/10.1007/978-1-4614-3223-4_2

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, *Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kadirhanoğulları, İ. H., Karadaş, K., Özger, Ö., Konu, M. (2021). Determination of Organic Product Consumer Preferences with Decision Tree Algorithms: Sample of Iğdır Province. *Yuzuncu Yıl University Journal of Agricultural Sciences*, *31*(1), 188-196. https://doi.org/10.29133/yyutbd.799465

Kaçmaz, K. S., Aşkan, E., & Topcu, Y. (2023). Consumer perception and purchase attitude towards genetically modified foods during the COVID-19 pandemic: The case of Erzurum, Türkiye. *Yuzuncu Yil University Journal of Agricultural Sciences,* 33(4), 543–555. https://doi.org/10.29133/yyutbd.1319800

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, *31*(1), 7-15. https://doi.org/10.1016/0020-0190(89)90102-6

Kao, A. ve Poteet, S. (2005). Text mining and natural language processing: Introduction for the special issue. *SIGKDD Explorations Newsletter*, *7*(1), 1–2. https://doi.org/10.1145/1089815.1089816.

Karami, A., Bennett, L. S., & He, X. (2018). Mining public opinion about economic issues: Twitter and the U.S. presidential election. *International Journal of Strategic Decision Sciences*, *9*(1), 11. https://doi.org/10.4018/IJSDS.2018010102.

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, *30*(3), 607-610. https://doi.org/10.1177/001316447003000308.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). Big data. The parable of Google Flu: traps in big data analysis. *Science (New York, N.Y.)*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human-Centered Informatics*, *5*(1), 1-167.

Loria, S. (2018). Textblob: Simplified text processing. https://textblob.readthedocs.io/en/dev/. Access date: 02.03.2025.

Maithili, K., Naveen Raja, S. M., Kumar, R. P., & Koli, S. (2023). A survey on NLP (natural language processing) and transactions on NNL (neural networks and learning systems). *E3S Web of Conferences, 430*, 01148. https://doi.org/10.1051/e3sconf/202343001148.

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qualitative Health Research*, *26*(13), 1753–1760. https://doi.org/10.1177/1049732315617444.

Moldovan, M.-G., Dabija, D.-C., Stanca, L., & Pocol, C. B. (2024). A qualitative study on the consumer behaviour related to food waste: romanian perspectives through word cloud and sentiment analysis. *Sustainability*, *16*(10), 4193. https://doi.org/10.3390/su16104193.

Nyariki, D.M. (2009). Household data collection for socio-economic research in agriculture: Approaches and challenges in developing countries. *Journal of Social Sciences*, *19*, 91 - 99.

OECD. (2023). AI in Policy Evaluation: Governing with Artificial Intelligence. OECD Publishing. Paris, https://doi.org/10.1787/795de142-en.

Paleologo, M., Castellini, G., Bosio, A. C., Fontana, M., & Graffigna, G. (2024). Exploring social media to understand perceptions of milk quality among farmers, processors, and citizen-consumers. *Foods*, *13*(16), 2526. https://doi.org/10.3390/foods13162526.

Parashar, S., Singh, S., & Sood, G. (2023). Examining the role of health consciousness, environmental awareness and intention on purchase of organic food: A moderated model of attitude. *Journal of Cleaner Production*, *386*, 135553. https://doi.org/10.1016/j.jclepro.2022.135553.

Pedersen, T. L. (2023). ggraph: an implementation of grammar of graphics for graphs and networks. R package version 2.1.1. https://CRAN.R-project.org/packag e=ggraph.

Penone, C., Giampietri, E. & Trestini, S. (2024). Exploring farmers' intention to adopt marketing contracts: empirical insights using the TOE framework. Agric Econ *12*, 39 (2024). https://doi.org/10.1186/s40100-024-00333-7.

Prasad, J. R., Prasad, R. S., & Kulkarni, U. V. (2008). A decision support system for agriculture using natural language processing (adss). In Proceedings of the International MultiConference of Engineers and Computer Scientists, *1*, 1-13.

R Core Team. (2024). R: a language and environment for statistical computing. R Foundation for Statistical Computing.

Rashid, A. B., & Kausik, M. A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*, *7*, 100277. https://doi.org/10.1016/j.hybadv.2024.100277.

Robinson, D. & Silge, J. (2022). *widyr: Widen, process, then re-tidy data* (R package version 0.1.5) Computer software. CRAN. https://doi.org/10.32614/CRAN.package.widyr.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53-65.

Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2018). Using text mining techniques for extracting information from research articles. In K. Shaalan, A. Hassanien, & F. Tolba (Eds.), Intelligent Natural Language Processing: Trends and Applications. Studies in Computational Intelligence, *Springer*, *740*, 235-250).. https://doi.org/10.1007/978-3-319-67056-0_18.

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing and Management*, *24*(5), 513-523.

Schnieder, M., Behbehani, R., Canhoto, A. I., Beltagui, A., Kumar, N., & McCloskey, J. (2024). Asking vs. listening: A comparative analysis of focus groups and text mining customer online reviews (Version 1). University of Sussex. https://hdl.handle.net/10779/uos.28174889.v1

Siegrist, J. (2012). Effort-reward imbalance at work: theory, measurement and evidence. Universitätsklinikum Düsseldorf, Institut für Medizinische Soziologie. https://www.uniklinik-duesseldorf.de/fileadmin/Fuer-Patienten-und-Besucher/Kliniken-Zentren-Institute/Institute/Institut_fuer_Medizinische_Soziologie/Dateien/ERI/ERI-Website.pdf.

Silge, J., & Robinson, D. (2016). Tidytext: Text mining and analysis using tidy data principles in R. *Journal of Open Source Software*, *1*(3), 37. https://doi.org/10.21105/joss.00037.

Soleimanian Gharehchopogh, F., & Abbasi Khalifelu, Z. (2011). *Evaluation of intelligent tutoring systems: Instruction and education approach*. *International Journal of Innovation, Management and Technology, 2*(5), 425–429.

Sørensen, M. P., Ravn, T., Bendtsen, A.-K., Reyes Elizondo, A., Kaltenbrunner, W., Ščepanović, R., Marušić, A., Kavouras, P., Labib, K., Tijdink, J. K., Veltri, G. A., & Bergmans, J. (2020). *Report on the results of the focus group interviews* (SOPs4RI Deliverable D5.2, Version 1.0). SOPs4RI Consortium. https://www.sops4ri.eu/wp-content/uploads/D5.2_Report-on-the-Results-of-the-Focus-Group-Interviews.pdf.

Stryker, C., & Holdsworth, J. (2025). Natural language processing: the future of AI. IBM. https://www.ibm.com/think/topics/natural-language-processing. Access date: 02.04.2025.

Surjandari, I., Naffisah, M. S. N., & Prawiradinata, M. I. (2015). Text mining of twitter data for public sentiment analysis of staple foods price changes. *Journal of Industrial and Intelligent Information, 3*(3), 253-257.

Taysi, M. R., İnci, H., & Karakaya, E. (2021). Determining organic product consumption preferences in terms of benefit, attitude and purchase intention. *ISPEC Journal of Agricultural Sciences, 5*(2), 463–475. https://doi.org/10.46291/ISPECJASvol5iss2pp463-475

Topaçan, Ü. (2016). Sentiment analysis in social media posts: A study on machine learning approaches (Doctoral dissertation). Marmara University, *Institute of Social Sciences*, Istanbul, Türkiye.

Turgut, H., Tokmak, İ., & Gucel, C. (2012). The effect of employees' organizational justice perceptions on their organizational commitment: a university sample. *International Journal of Business and Management Studies*, *4*(1), 21-30.

Viégas, F. B., & Wattenberg, M. (2008). Tag clouds and the case for vernacular visualization. *Interactions*, *15*(4), 49-52.

Vijayarani, S., & Janani, R. (2016). *Text mining: Open source tokenization tools – An analysis*. *Advanced Computational Intelligence: An International Journal (ACII)*, *3*(1), 37. https://doi.org/10.5121/acii.2016.3104.

Wang, Y., Ahmed, S., & Bee, A. (2023). Selective avoidance as a cognitive response: Examining the political use of social media and surveillance anxiety in avoidance behaviours. *Behaviour & Information Technology*, *43*, 1–15. https://doi.org/10.1080/0144929X.2023.2182609.

Whisner, C. M., Wang, H., Felix, S., & Maciejewski, R. (2016). Mining the twitter-sphere for consumer attitudes towards dairy. *The FASEB Journal*, *30*(1 Supplement), 897.2. https://doi.org/10.1096/fasebj.30.1_supplement.897.2.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686.

Ylilehto, M., Liljamo, P., Raatiniemi, L., & Kanste, O. (2021). Leader's perceptions of the impact of increasing use of e-health services on assessing the need for treatment: qualitative study on secondary care emergency department. *Finnish Journal of eHealth and eWelfare*, *13*(1), 32-48. https://doi.org/10.23996/fjhw.96161.

Yüksel, A. S., & Tan, F. G. (2018). Knowledge discovery in social networks using text mining techniques. *Journal of Engineering Sciences and Design*, *6*(2), 324-333. https://doi.org/10.21923/jesd.384791.

Yüzbaşıoğlu, R. (2021). Consumer preferences for cooperative branded products: the case of Tokat province central district. *Uluslararası Tarım ve Yaban Hayatı Bilimleri Dergisi*, *7*(3), 477-484. https://doi.org/10.24180/ijaws.943111.

Yüzbaşıoğlu, R., & Kaplan, E. (2019). Determination of environmental awareness in purchasing behavior of individuals (The Case of Tokat Central Country). *Rewieved Journal of Urban Culture and Management*, *12*(38), 214-224.