Turkish Journal of Engineering

https://dergipark.org.tr/en/pub/tuje

e-ISSN 2587-1366

Unveiling trends: a comprehensive analysis of state funded projects of Türkiye through content analysis and text mining

Talha Koruk*[1], Turgay Tugay Bilgin[2]

[1]*Bursa Technical University, Computer Engineering Department, Bursa Türkiye, talha.koruk@btu.edu.tr*
[2]*Bursa Technical University, Computer Engineering Department, Bursa Türkiye, turgay.bilgin@btu.edu.tr*

Cite this study: Koruk, T., & Bilgin, T. T. (2025). Unveiling Trends: A Comprehensive Analysis of State Funded Projects of Türkiye through Content Analysis and Text Mining. Turkish Journal of Engineering, Volume (Issue), 237-249.

https://doi.org/10.31127/tuje.1538068

Keywords
Text mining
N-grams
TF-IDF
PMI
TUBITAK project analysis

Research Article

Received:24.08.2024
Revised:04.08.2024
Accepted:08.10.2024
Published:01.04.2025

Abstract

This paper presents a comprehensive analysis of 12,724 projects approved by TUBITAK (The Scientific and Technological Research Council of Türkiye) between 2008 and 2022. The research employs advanced text mining techniques, including N-gram-based text categorization, TF-IDF, and PMI scores, to uncover trends in research and development activities in Türkiye. The analysis begins by examining the distribution of projects across years and regions, then focuses on five leading sectors: Information Technologies, Automotive, Machinery Manufacturing, Electrical-Electronics, and Defense Industry. The study identifies prominent themes and their evolution over time for each sector, thereby illuminating the dynamics of Türkiye's innovation ecosystem. The results highlight sector-specific trends as well as cross-sector common themes such as artificial intelligence, mobile applications, and sustainable technologies. This research provides valuable insights for policymakers, researchers, and industry stakeholders in shaping Türkiye's scientific and technological development. By leveraging text mining techniques on a large corpus of project data, the study offers a data-driven perspective on the changing landscape of innovation in Türkiye, contributing to a better understanding of national research priorities and emerging technological focus areas.

1. Introduction

Providing comprehensive trend analysis is important for identifying patterns which guiding research by identifying emerging areas of study and gaps in existing literature. Successful completion of project development tasks is closely linked to the level of support received, and it is important to communicate these needs assertively to ensure the success of the project. Projects require a heterogeneous network of resources, including financial aid, expert guidance, experiential knowledge, and avenues for potential sales. At the global level, a variety of councils are responsible for proposing, approving, and providing financial support for research initiatives. TUBITAK (The Scientific and Technological Research Council of Türkiye) represents a pivotal national funding agency within this landscape. On a global scale, various councils are responsible for proposing, approving, and providing financial support for research initiatives. In Türkiye, TUBITAK stands as a pivotal national funding agency within this landscape. Established in 1963 with the goal of enhancing Türkiye's scientific and technological capabilities, TUBITAK has consistently supported research and development (R&D) initiatives undertaken by Turkish researchers affiliated with universities, research institutions, and industry. The council's contributions have been instrumental in fostering scientific and technological progress within Türkiye, as evidenced by the substantial number of scholarly articles and patents generated by Turkish researchers on the global stage [1]. TUBITAK is a pivotal institution that offers a comprehensive platform for project success. With a spectrum of resources beyond financial assistance, TUBITAK serves as a guiding light in

Türkiye, cultivating a supportive environment for endeavours and nurturing ingenuity and scientific progress. The validation process carried out by TUBITAK confirms the value of projects and endorses them as nationally significant. These projects provide valuable insights, methodologies, and solutions that are integrated into the wider discourse of science and technology. TUBITAK examines project proposals to ensure they align with the council's objective of advancing Türkiye's scientific and technological boundaries. The approved projects contain valuable information for shaping Türkiye's future innovations. Understanding the range of projects approved by TUBITAK is essential for various stakeholders, including scholars, legislators, and business owners. These initiatives reflect the transformative path of the country's scientific and technological advancement.

This study applies text mining analysis to the vast repository of information from TUBITAK-approved projects, utilizing an N-gram-based text categorization method. This methodology forms the basis for extracting and analyzing meaningful insights from the texts of 12,724 TUBITAK projects across diverse business sectors, spanning five-year intervals from 2008 to 2022. This study aims to identify emerging concepts gaining prominence and established themes gradually fading away by tracking changes in TF-IDF or PMI scores over time. Identifying key concepts within the dataset is made possible by looking for high-frequency trigrams with increased TF-IDF and PMI scores. The dataset includes projects across various business sectors which some of them may be overrepresented or underrepresented, which could bias the results in terms of identifying emerging or fading themes. On the other hand, since all the projects are sourced from TUBITAK, there may be institutional or regional biases, reflecting the priorities or interests of TUBITAK rather than global trends.

2. Related work

An analysis of 2,323 TUBITAK 1001 projects executed between 2012 and 2020, utilizing social network analysis to investigate collaborative research communities. In this network, universities are represented as nodes interconnected by links signifying joint project collaborations. The resulting network comprises 193 universities (nodes) and 2,805 collaborative ties (links), involving a total of 8,205 researchers to examine scientific collaboration patterns and influence within the network [2].

The researchers concentrate on exploring themes and trends in various areas such as distance learning research. An analysis of 27,735 journal articles published between 2008 and 2018 is conducted, employing semantic content analysis with N-gram-based text categorization [3]. Similarly, exploring the years 2013-2022 and focusing on doctoral dissertations, the study reveals a prevalence of experimental and survey studies [4]. Another research aims to analyze scientific documents. The study introduces a text mining-based approach to categorize research articles into distinct subject categories, employing five methods for keyword extraction, including most frequent words, term frequency-inverse sentence frequency (TF-ISF), co-occurrence analysis, eccentricity-based extraction, and the textRank algorithm. To classify articles based on their extracted features, three classification algorithms—Naive Bayes, Random Forest, and Support Vector Machine (SVM)—are employed [5]. The text-mining method is utilized, analyzing 1090 theses and dissertations from Turkish universities, applying techniques such as keyword extraction, cluster analysis, and visualization. The impact of technology on education is underscored, emphasizing the necessity for educational technologies. Clusters are identified through text mining analysis, revealing three in study summaries and five in keywords [6].

In an alternative scholarly context, this research undertakes an examination of papers presented at the Computer and Instructional Technologies Symposium spanning the years 2007 to 2021, employing text mining as a methodological approach. The investigation meticulously reviews 3145 abstracts, culminating in the discernment of concept maps, thematic patterns, and the quantification of the interrelationships between concepts [7]. Furthermore, an analogous methodology is employed to scrutinize 574 postgraduate theses within the Department of Management Information Systems (MIS) from 2002 to 2020. Utilizing the Hidden Dirichlet Discrimination algorithm for text mining, the study forecasts the clustering of postgraduate theses into 11 distinct topics [8]. Employing a parallel approach, an analysis of banking literature spanning 2002 to 2013 is conducted through text mining and Latent Dirichlet Allocation, revealing credit risk management as the predominant trend. This study posits opportunities for further research aimed at bridging the divide between technical advancements and pragmatic applications within the banking sector [9]. Additionally, a bibliometric analysis delves into the scientific output of the science education research community in quantum physics over the period 2000 to 2021. Retrieving 1520 articles from reputable databases, the study delineates scientific production, preferred publication venues, key researchers, countries involved, and the overarching research topics [10]. The application of text mining techniques to scrutinize Mobile Language Learning (MLL) research from 2007 to 2016 is also highlighted, unveiling a discernible inclination toward English language learning. While existing studies predominantly concentrate on specific language skills like vocabulary and speaking, the imperative for additional research encompassing diverse languages, skills, and the sustained impact of mobile learning is advocated [11]. Further contributing to the discourse, a methodological roadmap for text mining is presented, elucidating the stages and techniques integral to the process, including Latent Dirichlet Allocation and Non-negative Matrix Factorization. The guide underscores practical considerations and culminates with exemplifications of text mining applications in realms such as social media analysis and document clustering [12]. Moreover, the introduction of the n-gram variables approach, an innovative text mining methodology implemented in Stata, is expounded upon. Emphasizing the treatment of sequences of words as distinctive features, this approach

is lauded for its heightened accuracy in tasks like classification and regression, reduced data sparsity, and enhanced interpretability. The potential applications of the n-gram variables approach are demonstrated through its utilization in analyzing immigrant visa interviews and patient narratives [13].

Recent research spans a diverse spectrum of topics, as evidenced by various studies. A comprehensive text mining analysis covering two decades uncovers a surge of interest in Virtual Reality (VR) and Augmented Reality (AR) within tourism, revealing advancements from virtual tours to immersive historical simulations, along with positive impacts on visitor satisfaction tempered by concerns about user acceptance and potential negative experiences [14]. In the organizational research domain, an article navigates through essential text preprocessing techniques, encompassing fundamental methods like tokenization, stop-word removal, stemming, and lemmatization, as well as advanced strategies such as part-of-speech tagging, named entity recognition, and sentiment analysis [15]. Shifting to academia, a study explores how academic libraries strategically utilize social media, emphasizing valuable insights derived from text mining for optimizing communication strategies, resource promotion, patron engagement, and knowledge sharing [16]. Furthermore, an innovative perspective emerges in the realm of cross-border e-commerce logistics, proposing a novel approach that integrates Kansei engineering with text-mining analysis of online content. This approach harnesses emotions identified from customer reviews and complaints to inform the design of specific service features, embodying a data-driven and customer-centric ethos [17]. Similarly, in the field of space communication, a recent review article highlights the potential of Inter-satellite Optical Wireless Communication (Is-OWC) to revolutionize data transmission by offering significantly higher data rates, larger bandwidth, and improved security compared to traditional radio frequency (RF) systems. However, challenges such as atmospheric effects, pointing and tracking, and space debris must be addressed. Current research focuses on advanced modulation techniques, adaptive optics, and photonic integrated circuits to overcome these obstacles and realize the full potential of Is-OWC [18]. Furthermore, a recent study analyzed the performance of the BeiDou Navigation Satellite System (BDS), a Chinese positioning service, by evaluating its position dilution of precision (PDOP), pseudorange multipath, and carrier phase signal-to-noise ratio (SNR). Results indicate that BDS performance is comparable to other Global Navigation Satellite Systems (GNSS), with improvements observed in precise point positioning (PPP) solutions from 2019 to 2023. These findings demonstrate the growing maturity and accuracy of BDS as a global positioning service [19].

The study of trends is a crucial aspect of research, as it helps to identify patterns and guide future investigations. Several studies have utilized trend analysis to understand various phenomena in Türkiye and beyond. For instance, one study examines long-term trends in temperature extremes across 42 weather stations in Türkiye from 1970 to 2018, focusing on the impact of urbanization on these trends [20]. Similarly, a detailed analysis of traffic safety trends across Türkiye's seven geographic regions over an 11-year period from 2006 to 2016 reveals that regional rankings in terms of traffic safety have remained stable, although specific safety measures show variability [21]. Another research effort utilizes an algorithm to develop a mathematical model that predicts monthly energy production, based on real measurement data collected between 2013 and 2018, demonstrating the application of trend analysis in energy forecasting [22].

Urbanization's impact on land use is another area where trend analysis proves valuable. A study investigating the effects of irregular urbanization on land use and land cover (LULC) in Kahramanmaraş province over 30 years highlights significant negative consequences, such as the reduction of fertile agricultural land and unplanned industrialization [23]. Additionally, trend analysis has been employed to examine climate data across various locations in Türkiye, utilizing statistical methods like Mann-Kendall, Spearman's Rho, and Innovative Trend Analysis to identify significant trends at a 95% confidence level [24]. Similarly, precipitation and temperature changes between the old climate period (1981-2010) and the new climate period (1991-2020) of Türkiye were examined based on 81 provinces and 25 water basins [25]. In a different context, the study of Air Pollutant Index (API) data from 2000 to 2019 in the Kuching region reveals trends in air pollution, particularly related to industrial activities and urbanization [26]. Similarly, another study applies artificial neural networks (ANN) and trend analysis to forecast sainfoin production in Türkiye from 1990 to 2020, using different models to predict production outcomes [27]. The application of neural networks in analyzing medical data trends further underscores the versatility of these methods in various research domains [28].

Collaboration is a key factor in scientific advancement, and studies employing social network analysis provide insights into these dynamics. One study, for example, investigates the trajectory of scientific collaboration within Spanish psychology from 1970 to 1989 by analyzing 2,891 articles published in 29 psychology journals. This research uses UCINET and NetDraw software to map out collaborative patterns among Spanish psychologists, highlighting the importance of social network analysis in understanding collaborative research [29]. Addressing a different subject, a study on tuberculosis trends from 2013 to 2018 employs a retrospective cross-sectional approach to analyze data within an institutional setting, filling a gap in the literature on tuberculosis trends [30]. Trend analysis is also applied in the assessment of industrial competitiveness. A comprehensive study of the Turkish air conditioning industry from 2001 to 2021 examines its international competitiveness using a variety of analytical tools, including Balassa's Revealed Comparative Advantages Index (RCA) and the Product Mapping method. This study represents a novel approach in evaluating the competitiveness of the industry, showcasing how trend analysis can be effectively utilized to understand market dynamics [31].

2.1. Our contribution

In this study, there is held remarkable analysis of 12,724 projects approved by TUBITAK between 2008 and 2022 by employing text mining techniques, including 3-gram frequency, TF-IDF, and PMI scores, to show leading industrial trends in research and development activities in Türkiye.

3. Method

In the digital age, much of the information is available in digital environments in text format. This encompasses a wide variety of textual data. Since there is huge textual data, it is stated that text mining is the most effective way to analyze and interpret these chunks; therefore, it automatically extracts meaningful information from large textual datasets [32]. For this reason, conceptual trends are revealed by analyzing the project information presented at TUBITAK using the N-gram text mining method. In this context, the formulated investigation involves a series of sequential procedures, which are outlined in Figure 1. This summarizes pipeline of text mining. The process starts with the collection of projects from TUBITAK, followed by the preprocessing stage, wherein the data is refined to eliminate unnecessary elements such as punctuation and special characters. Subsequently, word tokenisation divides the text into single words, and stop words are removed using a predefined stop list to eliminate common words. Also, NLP-based stemming reduces words to their root forms. The processed text is divided into filtered 3-grams, which capture groups of three consecutive words. These 3-grams are then subjected to analysis using statistical methods, including frequency, TF-IDF, and PMI. Finally, all values are combined and normalised to ensure consistency. This process enables the extraction of meaningful insights from large sets of text.

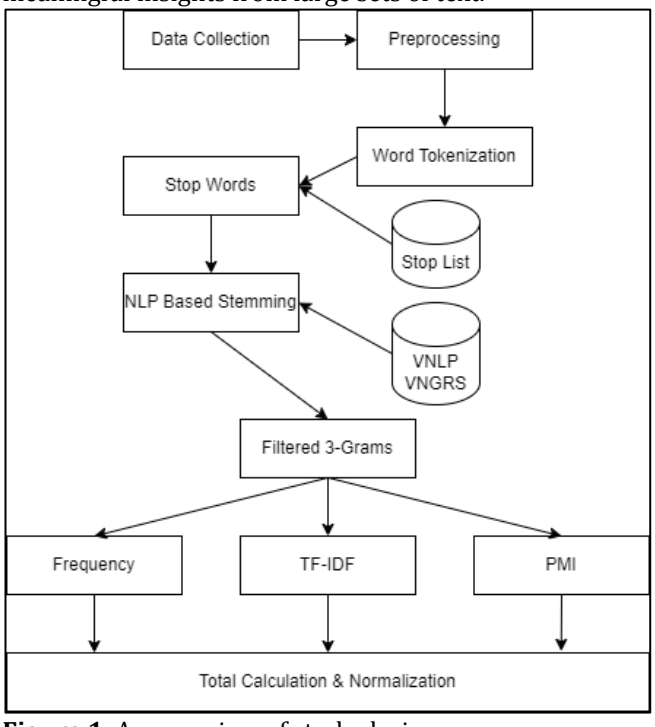


Figure 1. An overview of study design

3.1. Data collection

The rise of robotic process automation (RPA) has sparked a revolution in streamlining repetitive clerical tasks, empowering organizations to unlock new levels of efficiency. RPA bots, akin to digital apprentices, mimic human interactions with web and desktop applications, taking over tedious routines like file opening, form filling, and data entry [33,34]. This transformative technology finds its application in diverse settings, including the realm of project information management.

Considering the case of TUBITAK, their website houses a wealth of project information, crucial for research and development initiatives across the country. However, manually gathering this data often involves difficult navigation and repetitive tasks, hindering researchers and administrators alike. This is where RPA steps in as a game-changer.

The RPA tool is used to collect 22094 projects between 1995 and 2022. Each project includes information such as the project name, organization name, province of foundation, keywords, project date, scientific and technological activity, sector of execution, sector of output, project summary, purpose, and technical specifications.

3.2. Preprocessing

About half of the projects are missing important information such as abstract, purpose and technical specifications. Therefore, an N-gram analysis is conducted using 12724 complete project data from 2008 to 2022. Raw textual data from TUBITAK-approved projects undergoes text preprocessing through VNLP [35], a suite of Natural Language Processing (NLP) tools developed by the Turkish technology company VNGRS, caters specifically to the complexities of Turkish language processing. Its diverse functionalities, ranging from sentence splitting and normalization to stop-word removal, stemming, named entity recognition (NER), and sentiment analysis, provide valuable resources for researchers and practitioners alike. While the given information mentions the utilization of VNLP's stop-word cleaner and stemmer within a specific study, delving deeper into these tools and their academic potential warrants further exploration.

3.2.1. Word tokenization

The pre-processed text is tokenized, breaking it down into individual words. These elements, known as tokens, are grouped together as a semantic unit and used as input for further processing. This step is important for the stemming phase because VNLP uses AI-based functionality for stemming. Therefore, VNLP is looking for a full-period sentence to determine the stem of the word through context.

3.2.2. Stop words

The stop-word cleaner, a crucial element for text preprocessing, plays a significant role in NLP due to the language's agglutinative morphology and extensive

derivational processes. Understanding the specific stop-word list employed by VNLP and its approach to handling out-of-vocabulary words is crucial for evaluating its effectiveness in different applications. preprocessing to remove noise, irrelevant information, and standardize the format. This step ensures that the analysis is conducted on a clean and consistent dataset.

3.2.3. NLP based stemming

Stemming is a natural language processing (NLP) technique that involves reducing words to their base or root form, primarily by removing suffixes. This process helps in normalizing variations of a word to a common stem, thus minimizing the vocabulary size and enhancing text analysis. In this study, the VNLP stemming algorithm is employed to achieve this normalization.

3.2.4. Filtered 3-gram

The N-gram method, often used in text analysis, has several advantages that make it a valuable tool for understanding patterns and extracting information from textual data. By analyzing contiguous sequences of 'n' words, researchers can capture different levels of context. The frequency or presence of specific N-grams can be indicative of certain patterns or characteristics in the data. N-gram analysis can be considered well-suited for trend analysis over time, allowing researchers to track the evolution of patterns and thematic shifts within a dataset. In the context of TUBITAK-approved projects, the 3-gram (trigram) technique is used for the identification of trend themes and evolving patterns in the five prominent business sectors in Türkiye throughout the years.

3.2.5. Frequency

The frequency or presence of specific N-grams can be indicative of certain patterns or characteristics in the data. In the context of TUBITAK projects, the 3-gram frequency is the parameter that is used for the identification of trend themes and evolving patterns.

3.2.6. Term frequency-inverse document frequency

This technique involves using numerical statistics to assess the significance of a word within a collection of documents. TF-IDF, commonly employed as a weighting factor in information retrieval and text mining, reflects the importance of a word in a document by considering its frequency and presence across the collection. The TF-IDF value increases with the word's frequency in a document but is tempered by its prevalence across the entire collection. This method is useful for emphasizing fewer common words over more common ones, aiding in the control of document content. TF-IDF is particularly effective for filtering out stop-words in various subject fields, including text summarization and classification. Comprising two key statistics, term frequency (TF) and inverse document frequency, TF-IDF is calculated by evaluating how many times each term occurs in each document and summing these occurrences.

Equation (1) measures how often a trigram appears in each sector [36]. It is calculated as the ratio of the number of times a term occurs in a sector to the total number of trigrams in that sector.

$$TF(t,s) = \frac{Number\ of\ times\ t\ appears\ in\ s}{Total\ number\ of\ trigrams} \quad (1)$$

Equation (2) evaluates the importance of a term across a collection of sectors [36]. It is calculated as the logarithm of the ratio of the total number of sectors to the number of sectors containing the trigram, with the result inverted.

$$IDF(t,s) = log(\frac{Total\ number\ of\ s}{(Number\ of\ s\ include\ t)+1}) \quad (2)$$

In Equation (3) the TF-IDF score for a trigram in a sector is obtained by multiplying its Term Frequency (TF) by its Inverse Document Frequency (IDF) [36]. The higher the TF-IDF score for a term in a document, the more significant the term is in that sector relative to all sectors.

$$TF_{IDF(t,s,S)} = TF(t,s) * IDF(t,S) \quad (3)$$

3.2.7. Pointwise mutual information

Pointwise Mutual Information (PMI) is a statistical measure used to quantify the association between three terms (trigrams: w1, w2, w3), revealing the likelihood of their co-occurrence.

Equation (4) calculates the probability of the specific trigram (w1, w2, w3) occurring in the entire collection of trigrams [37]. It is the ratio of the number of occurrences of the trigram to the total number of trigrams.

$$P(w1,w2,w3) = \frac{Number\ of\ occurrence\ (w1,w2,w3)}{Total\ number\ of\ trigrams} \quad (4)$$

Equation (5), Equation (6) and Equation (7) show the calculation of the probability of each term occurring in trigrams [37]. It is calculated by dividing the number of occurrences of w1 in trigrams by the total number of trigrams. Also, this is applied for P(w2) and P(w3).

$$P(w1) = \frac{Number\ of\ occurrence\ w1\ in\ trigrams}{Total\ number\ of\ trigrams} \quad (5)$$

$$P(w2) = \frac{Number\ of\ occurrence\ w2\ in\ trigrams}{Total\ number\ of\ trigrams} \quad (6)$$

$$P(w3) = \frac{Number\ of\ occurrence\ w3\ in\ trigrams}{Total\ number\ of\ trigrams} \quad (7)$$

Equation (8) shows that PMI is computed by taking the logarithm of the ratio of the joint probability of the trigram w1, w2, w3 to the product of the individual

probabilities of its constituent terms (w1, w2, w3) [37]. This logarithmic measure reveals the deviation from independence. A positive PMI indicates a higher likelihood of co-occurrence, while a negative PMI suggests less likelihood than expected by chance. PMI is a valuable metric for capturing nuanced term associations beyond simple frequency-based measures.

$$PMI(w1, w2, w3) = log(\frac{P(w1,w2,w3)}{P(w1)*P(w2)*P(w3)}) \quad (8)$$

3.2.8. Total score calculation and normalization

In the context of term evaluation, a comprehensive total score is computed by combining three crucial factors: Term frequency, Term Frequency-Inverse Document Frequency and Pointwise Mutual Information. Equation (9) shows the total score calculation.

$$Score(t) = Freq(t) * TFIDF(t, s, S) * PMI(t) \quad (9)$$

To ensure comparability and eliminate biases due to varying scales, the total scores are normalized using the following Equation (10).

$$NormalizationOfScore = \frac{Score(t) - Min(s)}{Max(s) - Min(s)} \quad (10)$$

The total score calculation and subsequent normalization are essential steps in robustly assessing the significance of terms, enabling a fair and unbiased comparison of their importance within a specific sector and across a broader sector collection. As a result, normalized scores fit into “0.0” and “1.0”. Therefore, higher scores can be considered as valuable trigram terms.

4. Results

The primary objective of the initial analysis is to furnish a comprehensive description of the dataset. This encompasses an investigation into the yearly distributions and regional changes observed across five-year intervals. Following this, advanced 3-gram text processing methods, including the assessment of occurrence frequency, TF-IDF, and PMI scores, are employed. These analytical techniques are subsequently applied to conduct a thorough examination of five prominent business sectors: Information Technologies, Automotive, Machinery Manufacture, Electric Electronic, and Defense Industry.

Figure 2 shows the data distribution which represents the number of approved projects each year from 1995 to 2022. The project numbers show fluctuations over the years, with noticeable increases in the early 2000s, reaching a dip in 1999 and a peak in 2015. There is a subsequent decline, and the numbers stabilize around 1100-1300 projects per year.

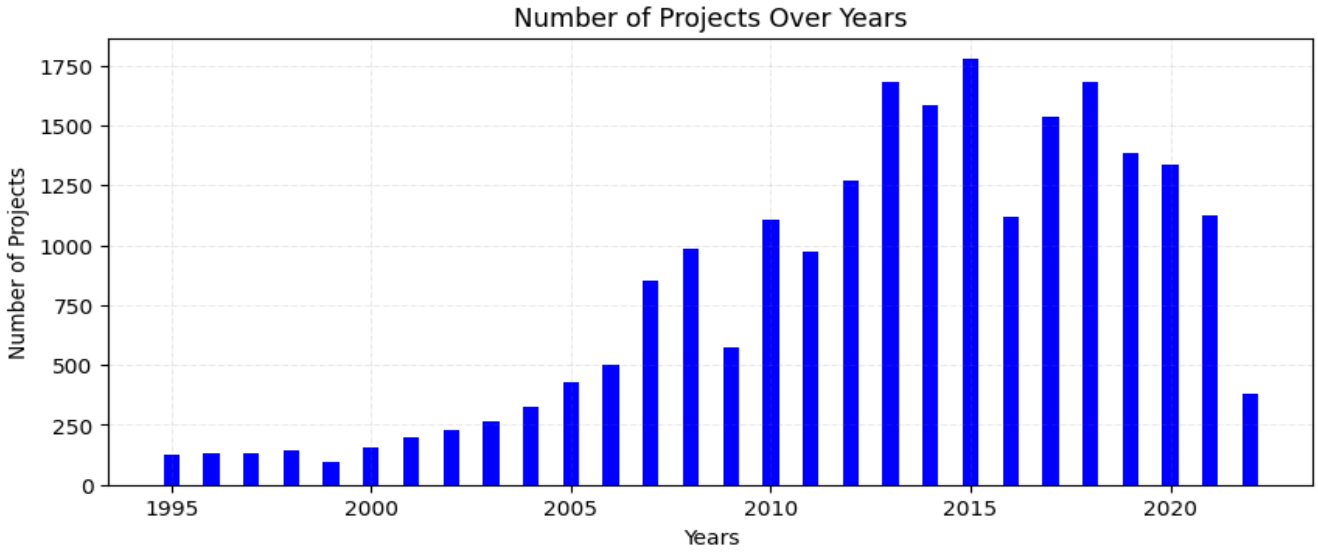


Figure 2. Distribution of the number of projects by years

First significant increase can be observed in 2007. The variations in project counts could be influenced by economic factors, technological advancements, changes in funding policies or unusual events. The peaks in the mid-2010s suggest a period of increased research and development activities. The decline post-2015 and subsequent stabilization may indicate a shift in project approval criteria or a maturation of ongoing projects. TUBITAK may consider analyzing these trends to adapt its support mechanisms or policies to align with the changing landscape of research and development in Türkiye. Focusing on understanding the factors contributing to peak years could provide insights into successful project strategies.

The given data displays the number of approved projects in different regions during three time periods (2008-2012, 2013-2017, and 2018-2022) as shown in Figure 3. Marmara stands out as the most prominent region, hosting a significant number of projects throughout all periods. Hence, Marmara can be regarded as the leading business region in Türkiye. Although

Marmara experienced an increase in the first and second time periods, it is also more vulnerable to unusual events than other regions. Aegean shows a notable increasing trend, indicating a growing emphasis on research and development activities. Mediterranean maintains a moderate but stable number of projects, while Black Sea shows a slight increase in project counts. Central Anatolian stands out with high project counts, particularly in the second period, indicating a significant surge. In contrast, both Eastern Anatolian and Southeastern Anatolian consistently have the lowest number of projects, indicating a possible need for increased attention or support in these regions. These observed disparities highlight the significance of regional balance in project distribution and urge policymakers to consider strategies that promote more equitable development. Furthermore, comprehending the factors that contribute to regional growth can aid in developing targeted initiatives to stimulate research and development activities where necessary. The periodic fluctuations in project counts highlight the dynamic nature of research endeavours and the need for adaptable policies to address evolving trends.

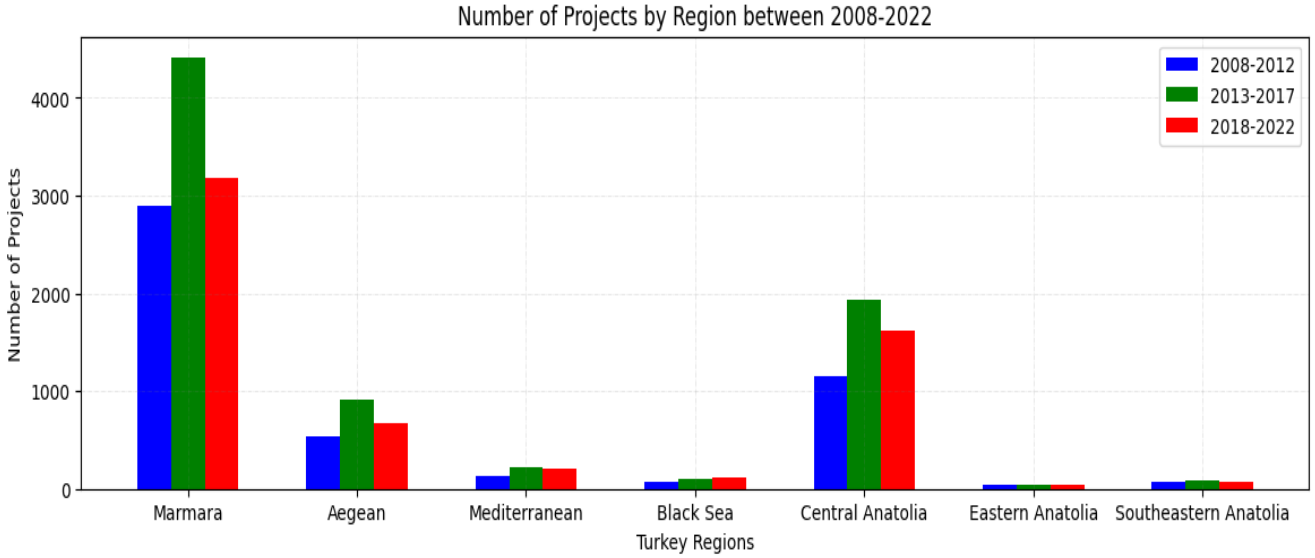


Figure 3. Distribution of the number of projects by region in 5-year periods

Table 1 presents an extensive summary of project distribution across major sectors, providing insights into sector-specific trends and dynamics over three five-year periods. The total number of projects increased from 2343 in 2008-2012 to 5328 in 2013-2017, followed by a slight decrease to 5053 in 2018-2022. The Information Technology, Automotive, Machinery Manufacturing, and Defence Industries have diverse trajectories that reflect changing priorities or market demands. Information Technology experienced a significant increase from 224 projects in 2008-2012 to 1037 in 2013-2017, followed by a slight decrease to 999 in 2018-2022.

Table 1. Distribution of the number of projects by top ten sectors in five-year periods

Sector Name	2008-2012	2013-2017	2018-2022
Information Tech.	224	1037	999
Automotive	483	792	499
Machinery Manuf.	273	662	433
Electronics	127	448	310
Energy	151	403	302
Biomedical	89	342	420
Defence Industries	134	282	310
Textile	128	268	264
Food	71	298	218
Pharmaceutical	126	237	206
Other Sectors	537	559	1092
Total Number	2343	5328	5053

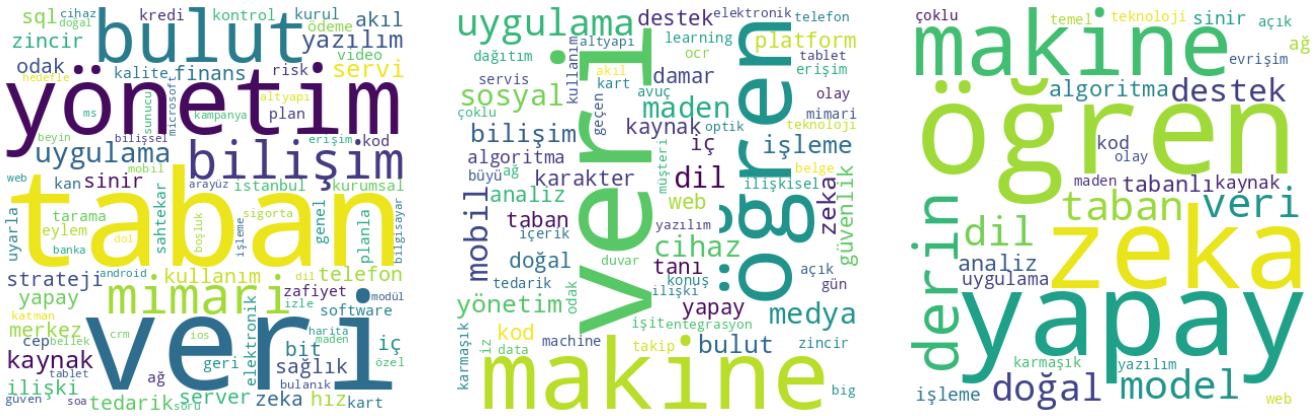
In the second period (2013-2017), Automotive had a peak of 792 projects, followed by a decrease to 499 in the third period. Machinery Manufacturing showed a consistent decline from 273 projects in 2008-2012 to 433 projects in 2018-2022. Defence Industries remained relatively stable, with a slight increase in the third period to 310 projects. Electronics and Energy exhibited fluctuations, possibly influenced by technological advancements or shifts in energy policies. The field of electronics saw a rise in projects over the first three periods, with a peak of 448 in 2013-2017, followed by a decline to 310 in 2018-2022. In the energy sector, there was an initial increase to 403 projects in 2013-2017, but this was followed by a decrease to 302 in 2018-2022. The

biomedical and pharmaceutical industries are experiencing substantial growth, with an increased focus on health-related research and development. The field of Biomedical research has shown a growth trend, with a peak of 420 projects in the period of 2018-2022. The number of pharmaceutical projects decreased from 237 in 2013-2017 to 206 in 2018-2022. The textile, food, and pharmaceutical sectors show varying patterns, reflecting the dynamic nature of consumer needs and industry trends. The number of textile projects decreased from 128 during 2008-2012 to 64 during 2018-2022. The number of projects in other sectors has significantly increased from 537 during 2008-2012 to 1092 during 2018-2022, indicating a diversification of project areas over time.

Table 2 presents the analysis of trigram frequency and five-year changes in the Information Technologies sector, highlighting dynamic trends and evolving priorities over the examined time periods (2008-2012, 2013-2017, 2018-2022). It is worth noting that trigrams such as 'kalite', 'kontrol', 'yazılım' and 'akıl', 'cep', 'telefon' consistently demonstrate high frequency, indicating sustained relevance. Recent years have seen an increase in the importance of certain trigrams, such as ('mobil', 'cihaz', 'uygulama') and ('web', 'uygulama', 'güvenlik'), indicating emerging trends. On the other hand, trigrams like ('web', 'taban', 'uygulama') and ('çoklu', 'dil', 'destek') have decreased in frequency, suggesting potential shifts in focus or decreasing significance. The analysis emphasizes the dynamic nature of the Information Technologies sector. Some trigrams continue to be prominent while others have become less relevant, reflecting the evolving landscape of technological priorities and advancements. The word cloud analysis presents an overview of the key themes and technological domains within Information Technologies, as depicted in Figure 4.

Table 2. Sector of information technologies trigram score and five-year change

2008-2012		2013-2017		2018-2022	
('kalite', 'kontrol', 'yazılım')	0.994	('mobil', 'cihaz', 'uygulama')	0.807	('yapay', 'zeka', 'makine')	0.857
('yapay', 'sinir', 'ağ')	0.993	('yapay', 'zeka', 'destek')	0.792	('derin', 'öğren', 'tabanlı')	0.672
('akıl', 'cep', 'telefon')	0.970	('platform', 'mobil', 'uygulama')	0.788	('web', 'taban', 'uygulama')	0.658
('kaynak', 'kod', 'yazılım')	0.967	('bulut', 'bilişim', 'teknoloji')	0.778	('çoklu', 'dil', 'destek')	0.655
('bulut', 'bilişim', 'mimari')	0.965	('web', 'uygulama', 'güvenlik')	0.768	('karmaşık', 'olay', 'işleme')	0.644



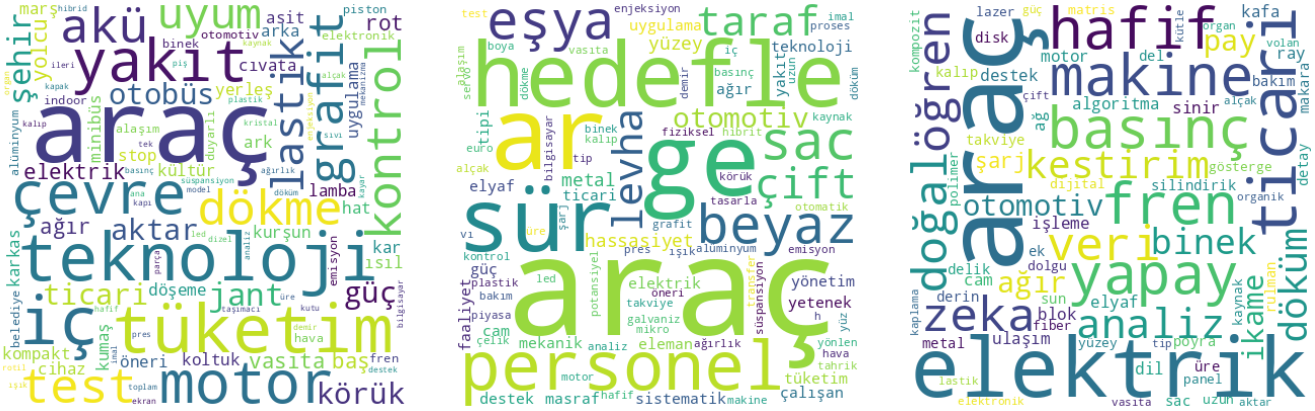
(a) 2008-2012 (b) 2013-2017 (c) 2018-2022

Figure 4. Sector of information technologies word cloud in 5-year periods. (a) Period of 2008 – 2012. (b) Period of 2013 – 2017. (c) Period of 2018 – 2022.

Table 3 presents an analysis of trigram frequency and five-year changes in the Automotive sector, revealing interesting patterns and dynamics. It is notable that trigrams such as ('kontrol', 'test', 'cihaz'), ('cam', 'elyaf', 'takviye'), ('yapay', 'zeka', 'destek'), and ('kurşun', 'asit', 'akü') consistently maintain high frequency across different time periods (2008-2012, 2013-2017, 2018-2022), indicating their enduring importance in the sector. The trigram ('ray', 'ulaşım', 'araç') had moderate frequency in the initial period but lacked subsequent data, suggesting potential variability or a changing emphasis over time. However, trigrams such as ('veri', 'analiz', 'kestirim'), show an increase in frequency in later years, indicating emerging trends or a heightened focus on data analysis and estimation within the automotive sector. This demonstrates the industry's growing interest and investment in data-driven decision making. The repetition of entries, such as ('cam', 'elyaf', 'takviye'), indicates different contexts or aspects related to the trigram. This analysis highlights the stability and evolution within the automotive sector. Some terms maintain their significance while others experience fluctuations, reflecting the dynamic nature of the sector and evolving technological priorities. Moreover, Figure 5 presents a word cloud analysis that provides a comprehensive summary of the main themes and technological domains in the automotive industry.

Table 3. Sector of automotive trigram score and five-year change

2008-2012		2013-2017		2018-2022	
('kontrol', 'test', 'cihaz')	0.992	('cam', 'elyaf', 'takviye')	0.988	('yapay', 'zeka', 'destek')	0.992
('kurşun', 'asit', 'akü')	0.984	('led', 'ışık', 'kaynak')	0.985	('ray', 'ulaşım', 'araç')	0.915
('koltuk', 'döşeme', 'kumaş')	0.979	('boya', 'galvaniz', 'sac')	0.967	('cam', 'elyaf', 'takviye')	0.902
('yakıt', 'tüketim', 'emisyon')	0.972	('mikro', 'alaşım', 'çelik')	0.966	('veri', 'analiz', 'kestirim')	0.894
('binek', 'ticari', 'araç')	0.970	('binek', 'tip', 'araç')	0.944	('binek', 'tip', 'araç')	0.842



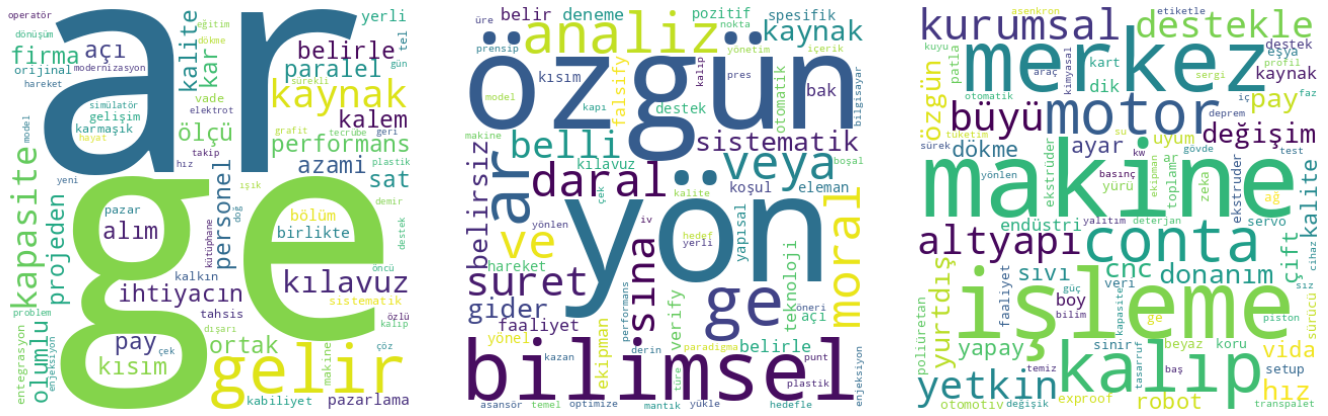
(a) 2008-2012 (b) 2013-2017 (c) 2018-2022

Figure 5. Sector of automotive word cloud in 5-year periods. (a) Period of 2008 – 2012. (b) Period of 2013 – 2017. (c) Period of 2018 – 2022.

The analysis of trigram frequency and five-year changes in the Machine Production sector, as depicted in Table 4, reveals notable trends and dynamics. Trigrams such as ('grafit', 'dökme', 'demir'), ('sistematik', 'mantık', 'yönetim'), ('karmaşık', 'ar-ge', 'entegrasyon'), and ('yerli', 'kaynak', 'üre') consistently exhibit high frequency, indicating their sustained significance in the sector over different time periods. Specifically, the trigram ('grafit', 'dökme', 'demir') reflects continued focus on materials science and metallurgical processes, which play a critical role in industries such as automotive and heavy machinery. The consistent appearance of this trigram highlights advancements in alloy composition and casting techniques. Similarly, the trigram ('sistematik', 'mantık', 'yönetim') points to a strong emphasis on systematic management practices, possibly related to the integration of modern software-driven methodologies in project and resource management. ('yerli', 'kaynak', 'üre') underscores the ongoing effort to enhance domestic production capabilities and reduce external dependencies. Additionally, trigrams like ('yapay', 'sinir', 'ağ') and ('servo', 'motor', 'sürücü') demonstrate moderate frequency, suggesting emerging trends with increasing significancy in recent years. The trigram ('yapay', 'sinir', 'ağ') reflects the growing importance of artificial neural networks in various applications such as real-time decision-making, pattern recognition, and machine learning-driven automation. This moderate but consistent frequency suggests that they are steadily gaining traction, likely in response to advances in edge computing and real-time analytics. On the other hand, the trigram ('servo', 'motor', 'sürücü') highlights developments in precision control systems, which are critical in robotics, automation, and industrial machinery. Its moderate frequency implies ongoing research and innovation in motion control technologies, perhaps driven by increased demand for high-precision manufacturing and smart factory solutions. Specific areas of emphasis are reflected in trigrams like ('operatör', 'eğitim', 'simülatör') and ('profil', 'işleme', 'makine'). However, the repetition of the trigram ('plastik', 'enjeksiyon', 'kalıp') raises questions about potential errors or differing contexts. Some trigrams, such as ('endüstriyel', 'bulaşık', 'makina'), exhibit low frequency, suggesting specialized applications or relatively lower importance. In essence, the analysis underscores a dynamic landscape within the Machine Production sector, encompassing both enduring themes and evolving priorities. Moreover, the word cloud analysis provides a comprehensive overview of the key themes and technological domains within the machine production in Figure 6.

Table 4. Sector of machine production trigram score and five-year change

2008-2012		2013-2017		2018-2022	
('grafit', 'dökme', 'demir')	0.911	('sistematik', 'mantık', 'yönetim')	0.967	('yapay', 'sinir', 'ağ')	0.758
('karmaşık', 'ar-ge', 'entegrasyon')	0.908	('yerli', 'kaynak', 'üre')	0.941	('servo', 'motor', 'sürücü')	0.705
('operatör', 'eğitim', 'simülatör')	0.759	('plastik', 'enjeksiyon', 'kalıp')	0.834	('poliüretan', 'sıvı', 'conta')	0.670
('geri', 'dönüşüm', 'makine')	0.717	('asansör', 'otomatik', 'kapı')	0.819	('profil', 'işleme', 'makine')	0.635
('plastik', 'enjeksiyon', 'kalıp')	0.291	('punt', 'kaynak', 'makine')	0.631	('endüstriyel', 'bulaşık', 'makina')	0.367



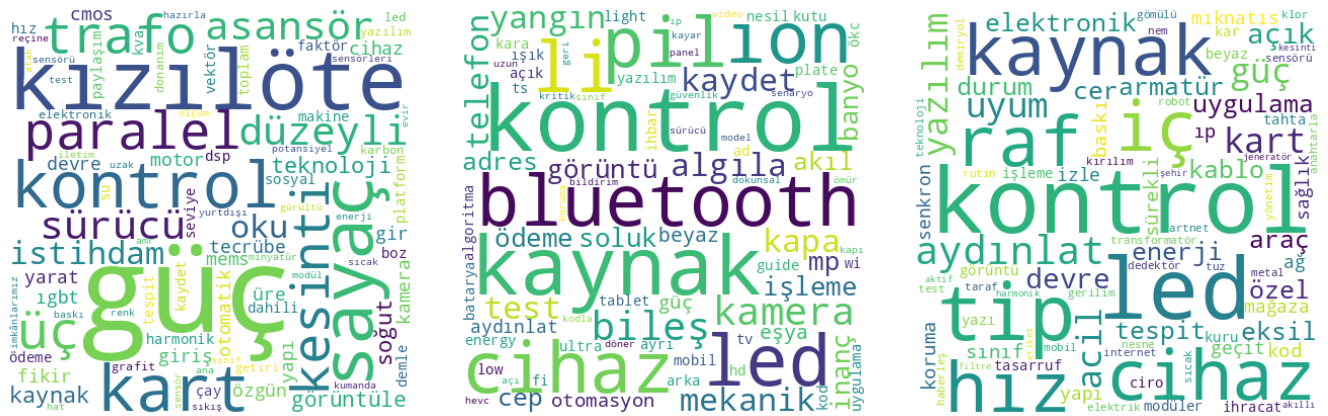
(a) 2008-2012 (b) 2013-2017 (c) 2018-2022

Figure 6. Sector of machine production word cloud in 5-year periods. (a) Period of 2008 – 2012. (b) Period of 2013 – 2017. (c) Period of 2018 – 2022.

The analysis of trigram frequency and five-year changes in the Electric Electronic sector, as delineated in Table 5, reveals intriguing patterns and dynamics. Notably, trigrams such as ('çay', 'demle', 'makine'), ('görüntü', 'işleme', 'algoritma'), ('motor', 'sürücü', 'asansör'), ('mobil', 'cihaz', 'uygulama'), ('ödeme', 'kaydet', 'cihaz'), ('beyaz', 'eşya', 'kontrol'), ('grafit', 'karbon', 'led'), and ('döner', 'kayar', 'kapı') consistently maintain high frequency across different time periods, underscoring their enduring importance in the Electric Electronic sector. Moreover, trigrams such as ('baskı', 'devre', 'kart'), ('görüntü', 'işleme', 'uygulama'), and ('robot', 'nesne', 'internet') exhibit moderate frequency in specific periods, suggesting the emergence of new trends or evolving technologies within the sector. Conversely, trigrams like ('kesinti', 'güç', 'kaynak') demonstrate low frequency, indicating potential specialized applications or reduced importance. The trigram ('modüler', 'güç', 'kaynak') displays variable importance with moderate frequency in one period, suggesting fluctuations in focus or relevance. This comprehensive analysis provides insights into the dynamic landscape of the Electric Electronic sector, encompassing both enduring themes and emerging trends, thereby contributing to a nuanced understanding of the sector's technological priorities and advancements over time. Moreover, the word cloud analysis provides a comprehensive overview of the key themes and technological domains within the electric electronics in Figure 7.

Table 5. Sector of electric electronic trigram score and five-year change

2008-2012		2013-2017		2018-2022	
('çay', 'demle', 'makine')	0.990	('görüntü', 'işleme', 'algoritma')	0.999	('baskı', 'devre', 'kart')	0.798
('motor', 'sürücü', 'asansör')	0.988	('mobil', 'cihaz', 'uygulama')	0.963	('görüntü', 'işleme', 'uygulama')	0.780
('ödeme', 'kaydet', 'cihaz')	0.986	('beyaz', 'eşya', 'kontrol')	0.906	('modüler', 'güç', 'kaynak')	0.779
('grafit', 'karbon', 'led')	0.951	('döner', 'kayar', 'kapı')	0.890	('robot', 'nesne', 'internet')	0.740
('kızılöte', 'kamera', 'modül')	0.926	('kesinti', 'güç', 'kaynak')	0.520	('gömülü', 'kart', 'yazılım')	0.722



(a) 2008-2012 (b) 2013-2017 (c) 2018-2022

Figure 7. Sector of electric electronic word cloud in 5-year periods. (a) Period of 2008 – 2012. (b) Period of 2013 – 2017. (c) Period of 2018 – 2022.

The analysis of trigram frequency and five-year changes in the Defense Industry sector, as outlined in Table 6, unveils a strategic technological landscape with consistent emphasis on certain trigrams. Notably, trigrams such as ('kimyasal', 'film', 'kaplama'), ('saldırı', 'tespit', 'önle'), ('konum', 'belirle', 'algoritma'), ('meteoroloji', 'hava', 'gözlem'), ('akustik', 'sinyal', 'işleme'), ('hava', 'araç', 'iha'), ('muharip', 'uçak', 'karakteristik'), ('zırh', 'muharebe', 'araç'), ('lazer', 'mesafe', 'ölçer'), ('karbon', 'fiber', 'destek'), ('zırh', 'araç', 'stanag'), ('platform', 'coğrafi', 'konum'), and ('mayın', 'eyp', 'dedektör') consistently exhibit high frequency across distinct time periods, underscoring their enduring significance in the Defence Industry. These trigrams span a range of applications, from chemical film coating and attack detection to location algorithms and aerial vehicle characteristics. Additionally, the inclusion of trigrams like ('füze', 'yükle', 'istif') and ('radar', 'dalga', 'sönümle') with moderate frequency indicates ongoing technological evolution and adaptability within the sector. Moreover, the presence of trigrams related to cutting-edge technologies, such as lasers ('lazer', 'mesafe', 'ölçer') and advanced materials like carbon fiber ('karbon', 'fiber', 'destek'), highlights a strategic focus on innovation and advanced capabilities. On the other hand, the focus on materials science, as evidenced by trigrams like ('karbon', 'fiber', 'destek') and ('zırh', 'araç', 'stanag'), suggests that research into advanced materials for better protection and durability remains a priority. These materials not only enhance the performance and survivability of military platforms but also contribute to reducing the weight of equipment, improving mobility, and ensuring compliance with international protection standards (such as STANAG). The presence of trigrams related to geographical positioning and mapping systems ('platform', 'coğrafi', 'konum') reflects the increasing use of geospatial technologies to enhance precision targeting. This analysis provides insights into the Defense Industry's engagement to the technological advancements and the strategic importance. Moreover, the word cloud analysis provides a comprehensive overview of the key themes and technological domains within the defense industry in Figure 8.

Table 6. Sector of defense industry trigram score and five-year change

2008-2012		2013-2017		2018-2022	
('kimyasal', 'film', 'kaplama')	0.963	('saldırı', 'tespit', 'önle')	0.900	('konum', 'belirle', 'algoritma')	0.967
('meteoroloji', 'hava', 'gözlem')	0.956	('akustik', 'sinyal', 'işleme')	0.899	('hava', 'araç', 'iha')	0.960
('muharip', 'uçak', 'karakteristik')	0.952	('zırh', 'muharebe', 'araç')	0.897	('lazer', 'mesafe', 'ölçer')	0.953
('karbon', 'fiber', 'destek')	0.947	('füze', 'yükle', 'istif')	0.805	('zırh', 'araç', 'stanag')	0.950
('platform', 'coğrafi', 'konum')	0.943	('radar', 'dalga', 'sönümle')	0.801	('mayın', 'eyp', 'dedektör')	0.929

(a) 2008-2012 (b) 2013-2017 (c) 2018-2022

Figure 8. Sector of defense industry word cloud in 5-year periods. (a) Period of 2008 – 2012. (b) Period of 2013 – 2017. (c) Period of 2018 – 2022.

5. Conclusion

This investigation utilizes N-gram-based text categorization to scrutinize the extensive TUBITAK-approved projects corpus. The aim is to unveil the latent structures and evolving trends shaping the innovation landscape in Türkiye. The analytical approach employed is a powerful tool for generating a nuanced comprehension of the recurrent themes and evolving patterns characterizing TUBITAK-approved projects over time. The expected results will provide valuable insights into the dynamic nature of scientific and technological innovation in Türkiye. This solid foundation will inform decision-making, drive evidence-based policy formulation, and align future research directions strategically. The findings have practical implications for startups, researchers, and industry stakeholders in Türkiye. They provide clear guidance on navigating project strategy and optimizing resource allocation, which will undoubtedly lead to a more efficient and impactful innovation ecosystem.

As a future work, a predictive model may be developed to forecast the success or impact of a TUBITAK-approved project based on principal themes (e.g., keywords, project goal, project summary and project technical details). Machine learning algorithms may be used such as regression, decision trees, or ensemble methods to build a predictive model. On the other hand, a recommendation system that suggests related TUBITAK-approved projects based on the topics or themes extracted from project descriptions may be built. Moreover, topic modelling techniques may be applied to identify key themes using project summary content. Furthermore, a comprehensive academic analysis can be considered as examining VNLP's integration into the study. Understanding the research question, methodology, and metrics employed in using the stop-word cleaner and stemmer would provide valuable insights into their performance and applicability. This information could then be used to compare VNLP with other Turkish NLP tools, explore its potential for real-world applications, and contribute to the advancement of Turkish NLP research.

Acknowledgement

This work would be enhanced by the presence of support.

Author contributions

Turgay Tugay Bilgin: The research gap of TUBITAK projects in literature and spell check**Talha Koruk:** Data collection from the website, carried out the experiments work and the theoretical calculations.

Conflicts of interest

The authors declare that they have no competing interests.

References

1. Kaska, O., Akin, H. K., Tokgoz, N., & Halicioglu, R. (2017). An Analysis of TUBITAK Projects' Budgets and Regional Distributions with Recommendations. *Journal of Current Researches on Social Sciences*, *7*(1), 59-66.
2. Unutulmaz, S. (2022). TÜBİTAK Projelerindeki Güçlü Araştırma İşbirliğinin Sosyal Ağ Analizi ile Dinamiklerinin Değerlendirilmesi. *Süleyman Demirel Üniversitesi Vizyoner Dergisi*, *13*(35), 810-828.
3. Gurcan, F., & Cagiltay, N. E. (2023). Research trends on distance learning: A text mining-based literature review from 2008 to 2018. Interactive Learning Environments, 31(2), 1007-1028.
4. TÖNGEL, E., AYDIN, A., Mehmet, K. A. R. A., & ÇAKIR, R. (2020). "Bilgisayar ve Öğretim Teknolojileri" ve "Eğitim Teknolojileri" Alanlarında Yazılan Yüksek Lisans ve Doktora Tezlerinin Araştırma Eğilimleri: 2013-2018 Döneminin Bir Görüntüsü. Ondokuz Mayis University Journal of Education Faculty, 39(1), 69-82.
5. Gürbüz, T., & Uluyol, Ç. (2023). Research article classification with text mining method. Concurrency and Computation: Practice and Experience, 35(1), e7437.
6. Kukul, V., & Aydin, K. (2021). Classification of the theses and dissertations in the field of computer education and instructional technology in Turkey: An investigation through text mining. Participatory Educational Research, 8(1), 279-291.
7. Erdoğdu, F., & Gökoğlu, S. (2022). Bilgisayar ve Öğretim Teknolojileri Alanına İlişkin Kavramsal Eğilimin Sempozyum Bildirileri Çerçevesinde Belirlenmesi: Metin Madenciliği Yöntemi. Uludağ Üniversitesi Eğitim Fakültesi Dergisi, 35(3), 601-622.
8. Çalli, L., Çalli, F., & Çalli, B. A. (2021). Yönetim bilişim sistemleri disiplininde hazırlanan lisansüstü tezlerin gizli dirichlet ayrımı algoritmasıyla konu modellemesi. Manas Sosyal Araştırmalar Dergisi, 10(4), 2355-2372.
9. Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. Expert Systems with Applications, 42(3), 1314-1324.
10. Bitzenbauer, P. (2021). Quantum physics education research over the last two decades: A bibliometric analysis. Education Sciences, 11(11), 699.
11. Hwang, G. J., & Fu, Q. K. (2019). Trends in the research design and application of mobile language learning: A review of 2007–2016 publications in selected SSCI journals. Interactive Learning Environments, 27(4), 567-581.
12. Karl, A., Wisnowski, J., & Rushing, W. H. (2015). A practical guide to text mining with topic extraction. Wiley Interdisciplinary Reviews: Computational Statistics, 7(5), 326-340.
13. Schonlau, M., Guenther, N., & Sucholutsky, I. (2017). Text mining with n-gram variables. The Stata Journal, 17(4), 866-881.
14. Loureiro, S. M. C., Guerreiro, J., & Ali, F. (2020). 20 years of research on virtual reality and augmented reality in tourism context: A text-mining approach. Tourism management, 77, 104028.
15. Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: Review and recommendations. Organizational Research Methods, 25(1), 114-146.
16. Al-Daihani, S. M., & Abrahams, A. (2016). A text mining analysis of academic libraries' tweets. The journal of academic librarianship, 42(2), 135-143.
17. Hsiao, Y. H., Chen, M. C., & Liao, W. C. (2017). Logistics service design for cross-border E-commerce using Kansei engineering with text-mining-based online content analysis. Telematics and Informatics, 34(4), 284-302.
18. Muataz Abdulwahid, M. ., Kurnaz, S., Kurnaz Türkben, A. ., Hayal, M. R., Elsayed, E. E. ., & Aslonqulovich Juraev, D. (2024). Inter-satellite optical wireless communication (Is-OWC) trends: a review, challenges and opportunities. Engineering Applications, 3(1), 1–15.

19. Akgül, V., Görmüş , K. S., Kutoğlu , Şenol H., & Jin, S. (2024). Performance analysis and kinematic test of the BeiDou Navigation Satellite System (BDS) over coastal waters of Türkiye. *Advanced Engineering Science*, *4*, 1–14.
20. Aykır, D., Atalay, İ., & Coşkun, M. (2022). Periodic Changes of Temperature Extremes at Some Selected Stations in Turkiye (1970-2018). *Coğrafya Dergisi*, (45), 69-83.
21. Özen, M. (2018). Comparative study of regional crash data in Turkey. *Turkish Journal of Engineering*, *2*(3), 113-118. https://doi.org/10.31127/tuje.385008
22. Yıldız, A. (2019). Predicting the energy production of a rooftop PV plant by using differential evolution algorithm. *Turkish Journal of Engineering*, *3*(3), 106-109. https://doi.org/10.31127/tuje.466953
23. Aliyazıcıoğlu, K., Beker, F., Topaloğlu, R. H., Bilgilioğlu, B. B., & Çömert, R. (2021). Temporal monitoring of land use/land cover change in Kahramanmaraş city. *Turkish Journal of Engineering*, *5*(3), 134-140. https://doi.org/10.31127/tuje.707156
24. Gündüz, F., & Zeybekoğlu, U. (2024). Analysis of temperature and precipitation series of Hirfanli Dam Basin by Mann Kendall, Spearman's Rho and Innovative Trend Analysis. *Turkish Journal of Engineering*, *8*(1), 11-19. https://doi.org/10.31127/tuje.1177522
25. Demirgül, T. ., Yılmaz, C. B. ., Zıpır, B. N. ., Kart, F. S. ., Pehriz, M. F. ., Demir, V. ., & Sevimli, M. F. . (2022). Investigation of Turkey's climate periods in terms of precipitation and temperature changes. *Engineering Applications*, *1*(1), 80–90.
26. Drahman, S. H., Maseri, H., Nap, M. C., & Hossen, Z. B. (2024). Twenty Years of Air Pollutant Index Trend Analysis in Kuching, Sarawak, Malaysia (2000-2019). *Sains Malaysiana*, *53*(3), 623-633.
27. Çelik, Ş., & Köleoğlu, N. (2022). Trend analizi ve yapay sinir ağları: Tarımda bir uygulaması. *Journal of Awareness*, *7*(1), 39-46.
28. Yao, Z., Chen, Y., Wang, J., Wu, S., Tu, Y., Zhao, M., & Zhang, L. (2021, December). Trend analysis neural networks for interpretable analysis of longitudinal data. In *2021 IEEE International Conference on Big Data (Big Data)* (pp. 6061-6063). IEEE.
29. Sala, F. G., Osca-Lluch, J., & Peñaranda-Ortega, M. (2021). Evolution of scientific collaboration within Spanish Psychology between 1970 and 1989. *Anales de psicología*, *37*(3), 589.
30. Negash, H., Legese, H., Adhanom, G., Mardu, F., Tesfay, K., Gebreslasie Gebremeskel, S., & Berhe, B. (2020). Six years trend analysis of tuberculosis in Northwestern Tigrai, Ethiopia; 2019: A retrospective study. *Infection and Drug Resistance*, 643-649.
31. İzgi, F., & Kavacık, M. (2024). Analyzing global competitiveness of Turkish air conditioning industry. *Turkish Journal of Engineering*, *8*(2), 209-234.
32. Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. Education and information technologies, 26, 205-240.
33. Leno, V., Polyvyanyy, A., Dumas, M., La Rosa, M., & Maggi, F. M. (2021). Robotic process mining: vision and challenges. Business & Information Systems Engineering, 63, 301-314.
34. Geyer-Klingeberg, J., Nakladal, J., Baldauf, F., & Veit, F. (2018). Process Mining and Robotic Process Automation: A Perfect Match. *BPM (Dissertation/Demos/Industry)*, *2196*, 124-131.
35. Vngrs-Ai. (n.d.). VNLP: State-of-the-art, lightweight NLP tools for Turkish language [Computer software]. GitHub. Retrieved from https://github.com/vngrs-ai/vnlp
36. Aizawa, A. (2003). An information-theoretic perspective of tf–idf measures. *Information Processing & Management, 39*(1), 45-65. https://doi.org/10.1016/S0306-4573(02)00021-3
37. Van de Cruys, T. (2011). Two multivariate generalizations of pointwise mutual information. In *Proceedings of the Workshop on Distributional Semantics and Compositionality* (pp. 16–20). Association for Computational Linguistics.