Topic Modeling of Popular Science Magazine Issues with Latent Dirichlet Allocation (LDA)

Munise Seckin - Kapucu
Eskisehir Osmangazi University

Yusuf Kartal
Eskisehir Osmangazi University

Abstract: TÜBİTAK is an institution that pioneers science in Turkey. In addition, it offers the opportunity to learn science at their level for various age groups with the journals it publishes. The scientific topics covered in the journals published by TÜBİTAK have evolved over the years, reflecting the dynamic nature of science. When examining literature studies on popular science, numerous studies are encountered. Popular science books and magazines are generally reviewed in these studies according to various subjects. However, studies that examine popular science magazines with the machine learning method are limited. It is essential to reveal how continuous and high-quality journals shape popular science in the context of projecting the popularity of future scientific topics. The Latent Dirichlet Allocation algorithm has been used intensively in topic modeling for many years. Within the scope of this study, multiple topics were identified for each journal by applying Latent Dirichlet Allocation to the textual content of the journals, and the distribution of each topic within the journal was calculated. As a result, this study automatically and independently extracted the topics and themes in the journals published by TÜBİTAK. These results allowed us to gain insight into how various content areas have evolved throughout the journal's history. With the results obtained from this study, new researchers can use the data to examine how topic change and/or topic expansion occur in a field.

Keywords: Latent Dirichlet Allocation, Natural language processing, Topic modeling, Topic detection

Introduction

Science is a tool developed by humanity to understand nature and explore the world around us. Scientific knowledge accumulates over time and is used to improve human life and enhance our understanding of the world. Therefore, science contributes to the advancement of society by combining both theoretical knowledge and practical solutions. When examining the nature of science, we encounter various ways of understanding it. In recent years, science has been defined through a postmodern perspective. According to this view, science is a subjective human endeavor that is dependent on theory and culture and is based on experimental observations (Schwartz, 2004). Along with this perspective, the characteristics of scientific knowledge on which scientists agree include its variability, subjectivity, creativity, and its influence by social and cultural contexts.

Scientific knowledge can change as a result of technological advancements, the emergence of new findings, socio-cultural shifts, or paradigm changes (Abd-El-Khalick, 2001). However, these scientific developments are not always evolutionary; sometimes they occur in a revolutionary manner. One of the most significant developments in science is the concept of a paradigm shift, proposed by Thomas Kuhn. A paradigm shift necessitates a comprehensive reevaluation of the existing scientific understanding. For example, major transformations such as Copernicus's heliocentric model of the universe, Newton's mechanical physics, and Einstein's theory of relativity have led to profound changes in the conventional understanding of science.

Scientific knowledge expands with every discovery. The influence of technology on these discoveries cannot be overlooked. Advances in fields such as genetic engineering, biotechnology, nanotechnology, and artificial intelligence have pushed the boundaries of science. As the boundaries of science expand and the amount of knowledge increases, the challenge of understanding this knowledge arises. Popular science emerges as a solution to this problem. It serves as a tool in making science accessible to the general public at a level they can understand. Popular science aims to make science more comprehensible, engaging, and trustworthy for a broader audience. In doing so, it fosters scientific curiosity, enhances scientific literacy, and bridges the gap between science and society.

Science can be conveyed to individuals in both school settings and outside of school through various media, including written, visual, and auditory forms such as books, magazines, radio, television, and the internet (Pramling & Säljö, 2007). These tools contribute to the socialization of science and can be examined within the scope of popular science publications. Popular science publications aim to appeal to a broader audience by addressing engaging topics and themes that are interconnected. Popular science publications are not limited to text alone; they also incorporate a wide variety of elements, such as images, diagrams, and various formats (Afonso & Gilbert, 2012). Popular science presents a vast amount of information through various formats, including videos, popular science books, magazines, and out-of-class learning environments—for example, science and art centers (Eroğlu, 2017).

One of the media through which science and popular science are conveyed to society is popular science magazines. These magazines serve as a communication channel between scientists and individuals not engaged in scientific fields, presenting information in everyday and understandable terms, often through a narrative or explanatory style (Stelnicki et al., 2011). Popular science magazines are important tools for creating scientific awareness and integrating science into the daily lives of society by presenting science in an entertaining and informative manner. Among the oldest and most respected popular science magazines in the world are *Scientific American, National Geographic, Popular Science, and New Scientist*. In Türkiye, TÜBİTAK (The Scientific and Technological Research Council of Türkiye) publishes various magazines to disseminate scientific knowledge and share scientific studies with the public. Additionally, by publishing scientific books and materials, TÜBİTAK helps spread scientific knowledge to society. TÜBİTAK's Popular Science Magazines—such as *Bilim ve Teknik* (Science and Technology), *Bilim Çocuk* (Science for Children), *Meraklı Minik* (Curious Little One), and *Bilim ve Genç* (Science and Youth)—make scientific information accessible to a broad audience. Scientific publications targeting young people, in particular, hold a significant place in the field.

Within the scope of this study, the *Bilim ve Teknik* (Science and Technology) magazine will be examined. Aimed at youth and adults, *Bilim ve Teknik* has been published monthly since 1967. The magazine covers scientific and technological topics in light of current developments. Its objectives include encouraging young people to pursue interests in scientific and technological fields, communicating scientific and technological work to the public, and connecting those interested in innovations in basic and applied sciences with the information they seek (TÜBİTAK, 2025). The magazine features articles on topics such as health, technology, the brain and mind, earth and life, space and physics, the environment and energy, and the social sciences.

When examining literature on popular science, numerous studies can be found. These studies typically focus on analyzing popular science books and magazines based on various topics (Aksüt, 2021; Çalık et al., 2021; Demircioğlu & Bilici, 2024; Dilli, 2024; Parkinson & Adendorff, 2004; Tatalovic, 2008). However, no studies have been encountered that examine popular science magazines using machine learning methods. Therefore, the content of popular science presented through TÜBİTAK has become a subject of interest within the scope of this study. This research aims to model the topics emphasized in popular science, in particular. Additionally, it analyzes the magazine content featured in popular science using artificial learning models. In this context, the study utilizes LDA (Latent Dirichlet Allocation) topic modeling.

Topic models are applied across various fields, including medical sciences, software engineering, geography, and political science (Jelodar et al., 2019). Topic modeling is one of the most potent techniques in text mining, used for data mining, hidden pattern discovery, and uncovering relationships between data and text documents (Jelodar et al., 2019). Although there are various methods for topic modeling, one of the most popular in this field is Latent Dirichlet Allocation (LDA). From a broad perspective, LDA-based topic modeling methods have been applied to information extraction in various fields, including natural language processing, text mining, and social media analysis. In topic modeling, a topic is a list of words that appear in statistically significant counts. A text can be an email, a book chapter, a blog post, a journal article, or any unstructured text (Jelodar et al., 2019). LDA represents topics according to word probabilities, and the words with the highest probabilities in each topic can usually provide a good indication of what the topic is (Jelodar et al., 2019).

Topic models have achieved considerable success in recent years and are applied in various fields, including the analysis of news articles, document tagging, topic-based search interfaces, and document summarization (Anupriya & Karpagavalli, 2015, January). Topic modeling not only helps a researcher identify popular topics or related fields in their area, but it can also be used to explore emerging topics or fields of study over time.

Literature

When the literature is reviewed, studies using topic modeling to examine trends in library and information science, scientific research, e-books, and journal abstracts can be found (Anupriya & Karpagavalli, 2015, January; Mohammed & Al-augby, 2020; Silwattananusarn & Kulkanjanapiban, 2022; Park & Song, 2013; Owa, 2021). In a study conducted by Anupriya and Karpagavalli (2015, January), the use of Latent Dirichlet Allocation was examined to suggest appropriate tags for journal abstracts. The abstracts were analyzed using LDA with CVB0 and Gibbs sampling techniques to determine the number of topics. The results showed that the extracted topics captured meaningful structure in the data and that effective topic models could be applied to tag journal abstracts into appropriate categories. In another study conducted by Park and Song (2013), the aim was to identify topic trends in the field of library and information science in Korea. The results showed that among these subfields, information science and library services were the two most focused areas. In a separate study conducted by Silwattananusarn and Kulkanjanapiban (2022), the evolution of information science research was examined through a bibliometric analysis and semantic mining approach. As a result of the study, 42,738 articles were extracted from Clarivate Analytics' Web of Science Core Collection (2010–2020) and analyzed within the scope of 10 main topics.

A study conducted by Mohammed and Al-Augby (2020) addressed a comparison study on full-text-based scientific unstructured text document classification (e-books), clustering words into a set of topics as key terms by applying the most popular topic modeling approaches, Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA). The results of this article showed that LDA produced better results than LSA. The best results from the LDA method achieved a coherence value of 0.592179 when the number of topics was 20, whereas the coherence value of LSA was 0.577302 when the number of topics was 10. In the study conducted by Owa (2021), the aim was to identify topics from scientific articles through topic modeling.

The Present Study

In this study, the aim is to model and extract the topics from articles published in TÜBİTAK's *Bilim ve Teknik* magazine. For this purpose, articles published in TÜBİTAK's *Bilim ve Teknik* magazine (between 2010 and 2024) were examined using LDA-based Topic Modeling, and trends in the research were identified.

Method

In this study, a natural language processing method called Latent Dirichlet Allocation (LDA) was used to analyze the articles published in TÜBİTAK's *Bilim ve Teknik* magazine (between the years 2010 and 2024) and to determine the topic distributions of text-based content. This method enables the automatic and independent extraction of topics and themes from published articles, allowing for the measurement of their prevalence in publications covered by *Bilim ve Teknik* over time. This measurement offers insight into how the topics featured in the magazine have evolved throughout its history. The dataset includes information such as year, month, issue, title, content, and the source internet address for each record. In the following sections, the details of the process are presented in depth.

Data Gathering and Preprocessing

Firstly, a dataset was created by compiling the open-access content of TÜBİTAK journals. This dataset includes the monthly published content from Bilim ve Teknik magazine, along with each piece's title, the magazine's month, year, and issue information. The internet source addresses from which the data were obtained are also recorded in this dataset on a content-by-content basis for each issue (Table 1). The dataset comprises 8,671 articles published in 180 issues of Bilim ve Teknik magazine between 2010 and 2024. The data were handled initially in Turkish, and no translation process was applied. Rows with missing content data were removed during the data preprocessing stage, and then the title and content columns were merged to create a single text field.

Table 1. Dataset features descriptions

Feature Name	Description
Year	Publication Year
Month	Publication Month
Issue	Issue No
Title	Title of the publication included in the journal
Content	Complete text content of the publication
Link	An online web address for accessing all other information

Preprocessing: Text Cleaning and Tokenization

Before being processed with LDA, the magazine contents underwent preprocessing. During this step, all letters in the text were first converted to lowercase, and words that do not convey meaning or do not significantly affect topic extraction—such as adverbs and conjunctions—were removed. Punctuation marks and numbers were also cleaned, leaving only words composed of alphabetic characters. Although there is no universally accepted word list for this purpose, the Natural Language Toolkit (NLTK) provides a Turkish stopword list that is commonly used in such contexts. In this study, the stopword list provided by the NLTK was utilized.

After data cleaning, the text contents were split into words, organizing each distinct word as a "token." To divide each content into word-level units based on sentences, the `word_tokenize()` function provided by the Natural Language Toolkit (NLTK) was used. This function is a widely used tokenization technique in natural language processing applications. Tokenization is the process of breaking down a text into smaller, manageable units. In this case, the text is tokenized at the word level—each word in a sentence is treated as a separate element and stored within a list.

Training the LDA Model

In this study, the Latent Dirichlet Allocation (LDA) model was trained using the Gensim library to perform topic modeling on text data. LDA is a generative probabilistic model that represents documents as random mixtures over latent topics, where a distribution over words characterizes each topic (Jelodar et al., 2019).

Creating a Dictionary and a Corpus

Using the tokenized texts, a dictionary and corpus were created. The Dictionary and doc2bow functions from the Gensim library were utilized for this purpose. The resulting dictionary contains a list of unique words appearing across all text contents. Based on this dictionary, each text was then transformed into a numerical form—specifically, into a bag-of-words (BoW) vector. This transformation allowed the textual data to be represented in a format suitable for topic modeling.

Determination of Model Parameters

One of the key parameters that directly affects the performance of the LDA model is the number of targeted topics. The number of topics can be determined through various methods in LDA. However, within the scope of this study, the topics presented by the *Bilim ve Teknik* magazine were taken into account, and the number of main topics offered—namely 44—was used as the num_topics parameter. To ensure the reproducibility of results, the random seed was fixed at 42. The entire document set was passed through the model 10 times. Additionally, the distribution between documents and topics was adjusted automatically. The parameters assigned for the LDA model are presented in Table 2.

Table 2. LDA Parameters and Values

Parameter	Value
num_topics	44
random_state	42
passes	10
alpha	auto
per_word_topics	True

Model Output

As a result of the model training, each topic was obtained along with its most dominant terms and their associated probability values. The resulting LDA model successfully identified the latent topics within the documents and provided topic labels that serve as a basis for time series analysis.

Mapping Keywords to Topics

A semantic matching method based on Sentence-BERT (SBERT) and cosine similarity was used to label the topics extracted by LDA automatically. For the SBERT model, sentence-transformers/paraphrase-multilingual-mpnet-base-v2, which yields strong results for Turkish texts, was used. At this stage, each of the 44 keyword groups and each topic label was converted into embedded vectors, and the semantic similarities between these vectors were calculated. Each keyword group generated by LDA was matched with the topic label to which it had the highest similarity, and this topic label was assigned as the topic for the corresponding content. The obtained matches were initially assigned automatically. The topics matched with the SBERT model were also reviewed by domain experts, and the accuracy of the SBERT-based similarity matching was qualitatively evaluated.

Topic Trends over Years

The predicted topic for each text content was combined with the corresponding year information, and topic frequency was calculated yearly. In this way, information was obtained on how frequently each topic was addressed annually.

Results and Discussion

Figure 1 presents the word clouds corresponding to the 44 distinct keyword groups generated via LDA. In these clouds, more frequently occurring words are represented in larger font sizes. These visualizations facilitate the interpretation of the inferred topics and help determine appropriate topic labels.

Figure 1. The LDA model generates keyword groups

The obtained word clouds were used in conjunction with the SBERT model to identify meaningful topic labels. The topic labels corresponding to the word clouds generated by the LDA model are presented in Table 3. The words included in this table are presented in their original form and the Turkish language, as they appear in the dataset.

Table 3. Keywords – topic mapping by SBERT*

Topic	Keywords
Food	food, agriculture, natural, plastic, important, oil, waste, nutrients, excess, according to
Climate	energy, sun, hydrogen, electricity, energy, carbon, fuel, obtained, heat, wind
Molecular Biology	dna, cell, carbon, protein, proteins, graphene, nan, three-dimensional, molecule, molecules
Aviation	athlete, athletes, salda, shoe, piezoelectric, herpes, to athletes, of the shoe, athletes, ski
Biotechnology	new, immunity, that it is, however, health, researchers, detected, revealed, more, then
Neuroscience	petri dish, in the society, hell, pylorus, ulcer, helicobacter, groups, in the dish, not needed
Marine Sciences	first, world, year, in the year, big, sea, happened, then, from, last
Technology	new, in a way, possesses, high, used, material, electronics, important, obtained, developed
Biology	footprint, fingerprint, ecological, forensic, sweat, traces, trace, traces, biometric
Physical	atom, light, energy, electron, laser, particle, high, wave, two, electricity
History	university, tübitak, national, Istanbul, first, observation, türkiye, observatory, year, line
Medicine	Professor, scientific, university, first, after, new, became, important, medicine, mathematics
Technology	virtual, super, reality, computer, augmented, computers, ibm, computers, real, processor
Neuroscience	neutrino, neutrinos, music, muon, neutrinos, tau, melatonin, middle, different, massage
Aviation	first, air, by, plane, turkish, mar, detected, then, in, year, possessed
Communication	bicycle, bike, gear, hypotheses, jury, hypothesis, medicine, you can add, to his bag, radio
Philosophy	özlem, çelik, ilay, kılıç, ekici, ocak, tuncay, bülent, gözcelioğlu, one each
Artificial Intelligence and Data B.	artificial, intelligence, new, human, robot, skillful, intuitive, learning, intellect, great
Philosophy	cave, of the cave, in the cave, bats, bat, of the bats, horses, caves, to the night, dolly's
Philosophy	tübitak, popular, books, publication, translation, printing, magazine, age, world, professor
Psychology	brain, brain, nerve, that, according to, researchers, sleep, done, cognitive, more
Robotics	foam, bracelet, bleeding, carrying, we introduced, robot, battery, stops, gb size, black box
Physical	nuclear, fusion, radiation, radioactive, uranium, neutron, reactor, atom, fission, thorium
Communication	language, face, different, for example, social, human, exist, knowledge, same, very
Philosophy	mirror, tesla, tesla's, botany, horticulture, garden, quartz, gokyiğit, plant
Philosophy	mont, works, you can listen, goldman, spider, on tv, spiders, channel, scissors
Philosophy	emir, dwarf, sport, dwarves, spider, cloud, ibrahim, mother, silver, azer
Food	center, insulin, kay, science, exhibition, centers, and, ser, sugar, visit
Technology	new, internet, exists, place, named, information, area, even, not, about
Materials Science	it happens, skin, what you're curious about, way, chemical, smell, provides, body, is found
Maths	number, two, different, question, number, example, first, month, solution, obtained
Geology	sea, türkiye, in our country, large, fauna, place, earthquake, nature, anatolia, nature
Technology	technology, research, oriented, tübitak, project, international, education, information, digital, within the scope of
Technology	smart, by, mobile, application, technology, phone, life, image, google, techno
Neuroscience	heart, cell, blood, tissue, stem, cells, bone, organ
Astronomy	sky, big, light, sun, star, stars, but, dark, mass, approximately
Materials Science	water, drinking, fabric, coffee, water's, fiber, purification, textile, to water, chlorine
Climate	however, researchers show that the climate is large, approximately, according to, high.
Philosophy	able to swim, filtering, in liquids, withstood, key, nan, two, however, high, then
Astronomy	near, moon, in the sky, heaven, day, mars, venus, hour, appearance, night
Philosophy	white, matte, two, there are, then, according to, black, last, does, solution
Animal Science	plant, nature, plants, species, invasive, alien, bird, food, types
Physical	quantum, two, however, classical, spl, in the form, theory, fundamental, that, knowledge
Space Technology	space, magnetic, earth, nasa, first, vehicle, moon, ground, world, satellite

*Scanned in Turkish.

It was observed that SBERT assigned the same topic labels to different keyword groups. Upon examining the keyword clusters generated by LDA, it was qualitatively evaluated that they correspond to the same topics. Therefore, these topic labels were merged. As a result, instead of the initially targeted 44 topic labels, 23 distinct topics were identified and matched with each article. The findings regarding these 23 topics, along with the number of contents associated with each, are presented in Table 4. According to the results obtained from LDA

and SBERT, the topics of Technology, Climate, Philosophy, Biotechnology, and Astronomy stand out as the most frequently addressed topics between 2010 and 2024, by a significant margin (Table 4).

Table 4. Article counts by topic

Topic(Original)	Topic(English)	Count
Teknoloji	Technology	2249
İklim	Climate	990
Felsefe	Philosophy	966
Biyoteknoloji	Biotechnology	836
Astronomi	Astronomy	705
Matematik	Mathematics	393
Tıp	Medicine	368
Jeoloji	Geology	295
Psikoloji	Psychology	278
Fizik	Physics	260
Tarih	History	217
Malzeme Bilimi	Materials Science	199
Uzay Teknolojisi	Space Technology	160
Yapay Zeka ve Veri Bilimi	Artificial Intelligence and Data Science	158
Gıda	Food	150
Deniz Bilimleri	Marine Science	100
Nörobilim	Neuroscience	95
İletişim	Communication	94
Havacılık	Aviation	88
Moleküler Biyoloji	Molecular Biology	38
Hayvan Bilimi	Animal Science	26
Haberleşme	Communication	4
Robotik	Robotics	2

Table 5. Topic counts per year distribution

Topic/Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Astronomy	56	50	38	44	43	44	46	55	51	57	48	43	56	45	29
Biotechnology	71	54	60	66	54	52	47	66	47	46	54	38	72	58	51
Marine Sciences	11	11	9	18	15	16	9	3	1	3	1	2	0	1	0
Philosophy	61	46	48	46	53	57	66	97	62	62	71	64	82	79	72
Physics	30	30	9	24	34	23	13	17	12	12	7	9	8	12	20
Food	12	17	13	13	20	8	13	4	10	6	3	11	7	7	6
Communication	0	0	1	1	2	0	0	0	0	0	0	0	0	0	0
Aviation	11	8	13	8	14	13	4	2	0	0	2	4	3	5	1
Animal Science	0	3	4	0	4	0	5	1	2	1	1	1	1	0	3
Geology	24	51	50	43	18	14	19	10	6	8	11	11	10	11	9
Materials Science	17	21	16	10	12	3	13	23	26	18	11	8	10	4	7
Mathematics	12	14	15	25	19	12	11	31	36	37	36	36	42	35	32
Molecular Biology	5	9	4	4	7	2	5	0	0	0	1	0	0	1	0
Neuroscience	20	6	12	12	12	7	3	1	3	4	3	0	4	2	6
Psychology	30	26	18	19	18	17	19	19	25	9	12	16	14	12	24
Robotics	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
History	20	10	15	9	17	20	13	14	13	17	13	13	13	15	15
Technology	205	180	145	222	254	267	239	152	104	76	96	89	78	74	68
Medicine	52	50	41	29	17	11	22	22	21	19	19	23	21	11	10
Space Technology	10	7	8	15	15	9	15	9	8	16	2	9	6	8	23
Artificial Intelligence and Data Science	3	3	3	3	6	6	12	16	16	15	14	4	15	18	24
Climate	52	48	39	67	81	71	83	64	59	59	60	78	85	71	73
Communication	9	15	9	12	4	6	5	5	12	5	3	3	2	3	1

The analysis of the distribution of these topics over the years, as identified by LDA, was conducted based on the frequency of the keywords within the corresponding contents. The data regarding this distribution is presented in Table 5. Based on the information in this table, it can be inferred that the most frequently mentioned topics have maintained their popularity over the years and have not shown significant variation over time.

The distribution of the topics listed in Table 5 by year is illustrated in Figure 2 as a line chart. As seen in Figure 2, Technology stands out as the most frequently covered topic in Bilim ve Teknik magazine articles over the years. Particularly after 2019, although there has been a decline in the coverage of the Technology topic, an increase in the popularity of its subtopics, such as Artificial Intelligence and Data Science, as well as Climate, can be observed (Figure 2).
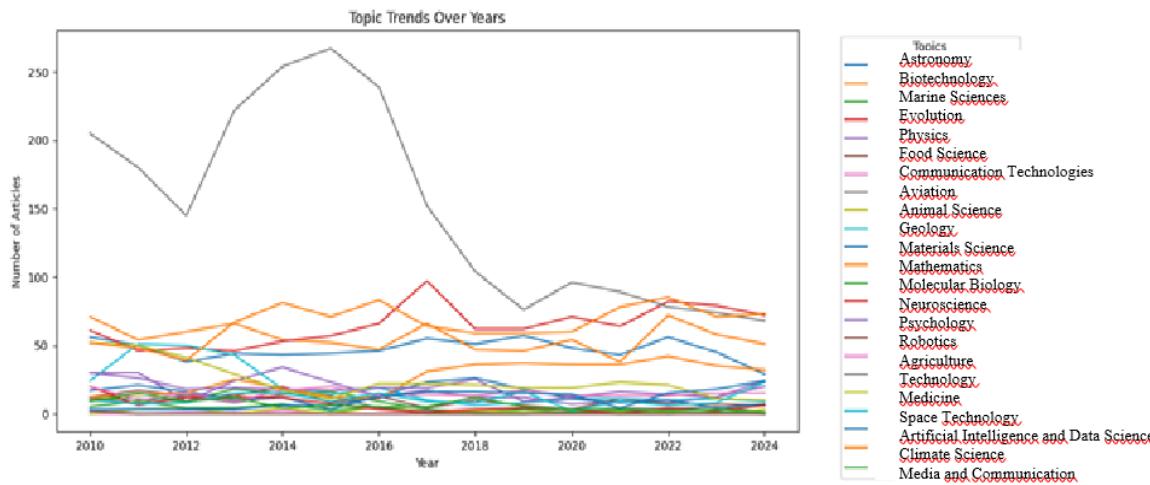

Figure 2. Topic trends over the years

The fact that *Bilim ve Teknik* is a technology-focused magazine makes the prominence of the Technology topic particularly meaningful. As a long-standing popular science magazine with a nationwide influence, the journal's editorial choices may explain why specific topics were covered even before they became widely popular. It should also be noted that, in line with its aim to address all branches of science, the magazine may have consistently included topics that, while not necessarily popular, reflect scientific continuity.

Conclusion

Since *Bilim ve Teknik* magazine aims to include content from every field of science, it can be considered that its texts broadly cover various scientific disciplines. In addition to topics that have been prominent on the public agenda for years, the magazine also addresses subjects that may not be widely discussed but are nonetheless studied by scientists. This suggests that the content encompasses a broad range of scientific disciplines.

In this study, the outputs of the LDA model were qualitatively evaluated through expert interpretation; therefore, no quantitative evaluation metric was used. As a result, the trained LDA model successfully identified latent topics across the documents and extracted keywords related to those topics, which serve as a foundation for time series analyses. In the models run by LDA and SBERT, the data with the highest matching scores were considered.

The findings indicate that, even without prior knowledge of any specific article in the *Bilim ve Teknik* magazine archive, it was possible to infer that the magazine generally emphasizes technological topics. Moreover, subfields such as Artificial Intelligence and Data Science emerged as core themes under the broader category of technology. It was also concluded that changes in interest toward specific topics over time can be effectively analyzed using an LDA-based natural language processing approach..

The SBERT model *sentence-transformers/paraphrase-multilingual-mpnet-base-v2*, which was used to identify the topics matching the extracted keywords, demonstrated notably high performance in topic prediction. This model, which is trained to capture semantic similarity across multiple languages, proved effective in aligning keyword sets generated by the LDA model with meaningful topic categories. Its multilingual structure also made it particularly suitable for processing Turkish texts, as used in the *Bilim ve Teknik* magazine content.

Recommendations

This study is limited to the data obtained from a restricted number of articles published in the *Bilim ve Teknik* magazine by TÜBİTAK. A larger dataset could be achieved by analyzing additional popular science magazines, which may lead to more robust results in practice.Expanding the stopword dictionary, experimenting with different parameters for the LDA model, or opting for language-specific models with a richer Turkish vocabulary, like those used in this study, could improve the accuracy of topic identification. The findings of this research may serve as a foundation for future researchers to investigate how topic shifts and/or topic diffusion occur within specific domains.

Scientific Ethics Declaration

* The authors declare that the scientific ethical and legal responsibility of this article published in EPESS journal belongs to the authors.

Conflict of Interest

* The authors declare that they have no conflicts of interest

References

Abd-El-Khalick, F. (2001). Embedding nature of science instruction in preservice elementary science courses: Abandoning scientism, but… *Journal of Science Teacher Education, 12*(3), 215–233.

Afonso, A. S., & Gilbert, J. K. (2012). The role of "popular" books in informal chemical education. *International Journal of Science Education, 34*(1), 77–99.

Aksüt, P. (2021). TÜBİTAK yayınlarından *Meraklı Minik* dergisinin fen kavramları açısından incelenmesi. *Journal of Individual Differences in Education, 3*(1), 1–19.

Anupriya, P., & Karpagavalli, S. (2015, January). LDA-based topic modeling of journal abstracts. In *2015 International Conference on Advanced Computing and Communication Systems* (pp. 1–5). IEEE.

Çalık, Ş., Koç, A., Şenel, T., Zor, E., & Aslan, O. (2021). Nanobilim ve nanoteknolojinin TÜBİTAK popüler bilim dergilerine yansımaları. *Uluslararası Türk Eğitim Bilimleri Dergisi, 9*(17), 90–113.

Demircioğlu, H., & Canbazoğlu Bilici, S. (2024). Popüler bilim dergilerinde mühendislik disiplininin ve mühendislik tasarım sürecinin incelenmesi: *Bilim Çocuk* dergisi örneği. *Anadolu Öğretmen Dergisi, 8*(2), 163–188.

Dilli, R. (2024). Tasarım odaklı düşünme yöntemi çerçevesinde popüler bilim dergileri içeriklerinin incelenmesi. *Journal of STEAM Education, 8*(1), 16–31.

Eroğlu, B. (2017). Fen eğitimi ve popüler bilim. In *Fen eğitimi üzerine güncel yaklaşımlar* (pp. 212–219). Ankara: Pegem Academy.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications, 78*, 15169–15211.

Mohammed, S. H., & Al-Augby, S. (2020). LSA & LDA topic modeling classification: Comparison study on e-books. *Indonesian Journal of Electrical Engineering and Computer Science, 19*(1), 353–362.

Owa, D. L. M. (2021). Identification of topics from scientific papers through topic modeling. *Open Journal of Applied Sciences, 10*(4), 541–552.

Park, J. H., & Song, M. (2013). A study on the research trends in library and information science in Korea using topic modeling. *Journal of the Korean Society for Information Management, 30*(1), 7–32.

Parkinson, J., & Adendorff, R. (2004). The use of popular science articles in teaching scientific literacy. *English for Specific Purposes, 23*(4), 379–396.

Pramling, N., & Säljö, R. (2007). Scientific knowledge, popularisation, and the use of metaphors: Modern genetics in popular science magazines. *Scandinavian Journal of Educational Research, 51*(3), 275–295.

Schwartz, R. S. (2004). *Epistemological views in authentic science practice: A cross-discipline comparison of scientists' views of nature of science and scientific inquiry* (Unpublished doctoral dissertation). Oregon State University, Corvallis, OR, USA.

Silwattananusarn, T., & Kulkanjanapiban, P. (2022). A text mining and topic modeling based bibliometric exploration of information science research. *IAES International Journal of Artificial Intelligence, 11*(3), 1057–1066.

Tatalovic, M. (2008). Student science publishing: An exploratory study of undergraduate science research journals and popular science magazines in the US and Europe. *Journal of Science Communication, 7*(3), A03.

TÜBİTAK. (2025). *Bilim ve Teknik*. Retrieved from https://bilimteknik.tubitak.gov.tr/dergimiz/hakkimizda

Author(s) Information

Munise Seckin -Kapucu	**Yusuf Kartal**
Eskisehir Osmangazi University, Faculty of Education, Mathematics and Science Education Department, Eskisehir, Türkiye. Contact e-mail: *muniseseckin@hotmail.com*	Eskisehir Osmangazi University, Faculty of Engineering and Architecture, Computer Engineering Department, Eskisehir, Türkiye

To cite this article:

Seckin-Kapucu, M. & Kartal, Y. (2025). Topic modeling of popular science magazine issues with Latent Dirichlet Allocation (LDA). *The Eurasia Proceedings of Educational and Social Sciences (EPESS), 47,* 57-66.