

Topic Analysis of Published Theses in Obstetrics and Gynecology

Ayşe KONAÇ^{1*} 

¹Assistant Professor, MD, Gelisim University, Faculty of Health Sciences, Department of Obstetric and Gynecology İstanbul, Türkiye

<p>Abstract: Objective: The objective of this research is to perform a thorough analysis of the topics covered in Obstetrics and Gynecology theses and determine the most frequently researched areas and subjects of specialization theses. The study aims to provide insights into the future direction of Obstetrics and Gynecology specialization theses by identifying emerging trends and patterns in previous research. Additionally, the research aims to contribute to the advancement of knowledge in the field by identifying gaps and limitations in the current literature and providing a direction for future research to improve patient care. The study's comprehensive analysis of these topics and identification of emerging trends and patterns contribute to the development of Obstetrics and Gynecology as an important field of study. Methods: The methodology of this study involved a comprehensive review of 4631 studies published in the field of "Obstetrics and Gynecology". The analysis was conducted using the Python (Python, 2022) programming language, which provided a flexible and efficient means of analyzing large volumes of data. Results: This study analyzed a subset of the academic theses stored in Turkey's National Thesis Center (UTM) related to Obstetrics and Gynecology. Out of the total 759,771 registered studies published since 1959, only 4631 relevant theses were analyzed. The study aimed to identify the most frequently researched topics and emerging trends in the field, which could provide insights into the future direction of Obstetrics and Gynecology specialization theses. The paper presents the detailed results of the analysis and offers a valuable contribution to the field of Obstetrics and Gynecology. Conclusion: Expertise theses in the field of Obstetrics and Gynecology in Turkey have been analyzed using topic analysis. The results have shown that the theses have focused on obstetric and gynecological diseases, infertility, obstetric and gynecological surgery, cancers, pregnancy monitoring, childbirth management, perinatology, prenatal diagnosis, and ultrasound. It is recommended that future expertise theses cover broader topics and that studies addressing psychological and social factors should also be conducted.</p>	<p>Research Paper</p>
	<p>*Corresponding Author: Ayşe KONAÇ Assistant Professor, MD, Gelisim University, Faculty of Health Sciences, Department of Obstetric and Gynecology İstanbul, Türkiye</p>
	<p>How to cite this paper: Ayşe KONAÇ (2026). Topic Analysis of Published Theses in Obstetrics and Gynecology. <i>Middle East Res J. Med. Sci.</i>, 6(2): 67-77.</p>
	<p>Article History: Submit: 28.12.2025 Accepted: 29.01.2026 Published: 07.03.2026 </p>
<p>Keywords: YOK Thesis, Bibliographic Databases, Gynecology and Obstetric, Evidence-Based Medicine, Internet, Python Program, Social Network Analysis.</p>	
<p>Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY-NC 4.0) which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.</p>	

1. INTRODUCTION

Obstetrics and gynecology is an important medical specialty that deals with women's reproductive health. This field is of great importance not only for individuals' health but also for public health. Women need obstetric and gynecological services during important periods of their lives such as pregnancy and childbirth. Additionally, these services play a crucial role in preserving and improving women's health during different periods of their lives.

Obstetric and gynecological services are critical for both maternal and fetal health. During the prenatal, intrapartum, and postpartum periods, these services help to preserve and improve the health of women and babies.

Additionally, gynecological services help to preserve women's reproductive health and aid in the early diagnosis and treatment of diseases. There is an academic debate about the potential mental health problems that may arise from inadequately provided obstetric and gynecological services for maternal and fetal health [1].

It is necessary to continue the interest and investments in this field to improve women's quality of life, preserve their health, and improve their reproductive health. Therefore, there is a need to analyze the literature in this field, giving particular importance to the analysis of the theses, which are the most comprehensive and detailed studies of obstetric and gynecological research.

Theses are considered an important resource for determining and developing the direction of future research. Therefore, the analysis of theses will provide guidance for those who want to conduct research in this field and help identify gaps in the literature, which will enable us to answer important questions for future research.

The field of obstetrics and gynecology is continuously updated with the latest research findings. In a time of rapidly published studies, there are several difficulties in interpreting the literature on obstetrics and gynecology [2]. To overcome these difficulties, research in obstetrics and gynecology must be analyzed correctly. This can only be achieved through a proper evaluation and interpretation of previous studies.

Today, with the widespread use of the internet and easy access to databases worldwide, it has become easier to access sources online when writing theses and articles. For this purpose, it is possible to access all theses via sources such as WOS, MedLine, Pub Med (NLM), and the YÖK Thesis database in Turkey, as well as many other databases. Especially in medical education, these databases are used as important sources of information. Through these databases, students and researchers can easily access up-to-date and reliable sources of information on their topics. This increases the quality of research and learning in medical education.

The Publication and Documentation Department of the Council of Higher Education, which was restructured as the "National Thesis Center of YÖK" in 1996, is a center responsible for collecting, organizing, and providing electronic access to theses accepted electronically since 2006. It is affiliated with the Presidency of the Council of Higher Education and provides all its services online. Graduate theses prepared at universities in Turkey are presented to internet users in digital form through the Electronic Thesis Archive Project, which began in 2007. As of 2022, the National Thesis Center archive contains almost 700,000 theses, and about two-thirds of them are available in full text [3].

This study aims to perform topic analysis of the theses published in the field of Obstetrics and Gynecology in Turkey. Working on 4631 theses registered in the National Thesis Center, analyses were performed using the Python programming language. The obtained data was examined using exploratory data analysis techniques, and then the topic modeling process was carried out using the Gensim library and the LDA algorithm.

The results of this study are expected to help determine the topics on which research on women's

health is focused in Turkey and identify areas where more work is needed in the future.

2. METHODOLOGY

This study aims to conduct a topic analysis of the theses published in the field of Obstetrics and Gynecology, and to determine the main topics covered in these specialized theses. Another objective of this study is to provide a perspective for future specialized theses.

In Turkey, academic theses are digitally stored in the National Thesis Center (UTM) [4], and are published as open source for use in academic studies. As of January 07, 2023, a query conducted in the UTM revealed that there were 759,771 registered studies since 1959. Only 4631 studies conducted in the field of Obstetrics and Gynecology, which are the subject of this study, were taken into account.

All analyses in this study were performed using the Python programming language [5].

2.1. Statistical Analysis

The analysis involved several steps: Firstly, the search strategy was set to "Subject=Obstetrics and Gynecology" and "Year=2000-2022". Secondly, only studies with English abstracts were included. Thirdly, exploratory data analysis examined various aspects of the published studies. Finally, descriptive statistics were provided for all Obstetrics and Gynecology theses, with specialized theses undergoing further topic analysis.

2.2. Study Strategy

The obtained studies were subjected to analysis by following the processes listed below.

1. Determination of search strategy: The search strategy was determined as "Subject=Obstetrics and Gynecology" and "Year=2000-2022".
2. Determination of language criteria: Only studies with English abstracts were included in the study.
3. Conducting exploratory data analysis: Year of publication, thesis title, thesis type, topic, university name, and abstract sections of the published studies were examined.
4. Analysis: Descriptive statistics of all theses conducted in the field of Obstetrics and Gynecology were given within the scope of this study, while only specialized theses were included in the topic analysis.

2.3. Topic Model

Gensim is an open-source document-based learning library written in Python. This library allows various document-based learning techniques (e.g. Latent Semantic Analysis (LSA), Latent Dirichlet Allocation

(LDA), and Hierarchical Dirichlet Process (HDP)) to be performed, enabling tasks such as document classification, similarity measurement, and topic model creation [6].

Topic modeling is a learning technique that aims to discover specific topics in a document collection. In the Gensim library, topic models can be created using an algorithm called LDA (Latent Dirichlet Allocation). LDA is a generative probability model that assumes each topic is a mixture of underlying word sets and each document is a mixture of topic probabilities [7, 8]. This algorithm assigns documents to specific topics based on the words found in a document collection and measures how important these topics are. This method helps to explore the content of a document collection and can be used for operations such as document classification [9, 10].

In this study, the Gensim library and LDA algorithm were used. Topic visualization was carried out using the pyLDAvis visualization package [11-12].

3. RESULTS

This section presents the descriptive analyses and Topic Model analysis findings conducted using the Python programming language in the scope of the study.

3.1. Descriptive Statistics

The distribution of the theses according to their types is shown in Figure 1 within the scope of the topic.

Accordingly, the dataset includes a total of 3,344 specialization theses in medicine, 996 master's theses, 275 doctoral theses, and 16 subspecialty theses in medicine. The distribution of theses by year for the topic area is shown in Figure 2.

The dataset covers the years 2000 to 2022. Accordingly, the highest thesis production, 464 theses, was realized in 2019. The highest number of specialization theses were produced in 2022, with 330, and the lowest number was 47 in 2001. Theses produced as subspecialties were carried out only between 2013-2020 (excluding 2014).

The distribution of the top 10 universities with the most theses produced on Obstetrics and Gynecology (specialization theses and others) is shown in Figure 3.

Accordingly, the institution that produced the most theses was the University of Health Sciences with 801 theses, followed by the Ministry of Health with 320 theses and Istanbul University with 257 theses. The distribution of the top 10 universities with the highest number of specialization theses produced is shown in Figure 4.

According to the analysis, the highest number of specialization theses was produced at the University of Health Sciences (771 theses), followed by 318 theses from the Ministry of Health units and 178 theses from Istanbul University. The distribution of specialization theses and other types of theses by year for the topic area is shown in Figure 5.

Accordingly, while the number of specialization theses and other types of theses had a partially parallel linear relationship until 2017, there has been an increase in the number of other types of theses since then.

The word cloud consisting of the 250 most frequently occurring words in the corpus obtained by tokenizing specialization theses is shown in Figure 6.

3.2. Topic Model, Coherence Score, and Visualization

In this study, the Gensim library was used to obtain topics from the corpus, and the pyLDAvis library was used to visualize the similarities between the topics.

Coherence Score

Different algorithms (C_v , C_p , C_{uci} , C_{umass}) and the Gensim library can be used to determine the number of topics to be obtained from the corpus (8,13). In this study, the coherence analysis was performed using the ' c_v ' algorithm to determine the number of topics. The values for the topic coherence scores obtained from the evaluation are shown in Figure 7.

In this study, it is observed that the scores obtained by the C_v coherence algorithm reach up to 0.70. It is known that a coherence analysis score above 0.50 is considered a good score (14). Therefore, considering this, the number of topics to be obtained from the dataset is determined as five (Coherence coefficient=0.673).

Visualization of Topics

To visualize the obtained topics, the pyLDAvis graph is used to reveal similarities and facilitate interpretation (12,15). This graph has three important features. Firstly, each circle represents a topic. The size of the circle indicates the prevalence of that topic, i.e., the larger the circle, the more studies have been conducted on that topic. Secondly, the distance between two circles indicates the similarity between topics. If the circles are far from each other, it shows that the topics are different. Finally, on the right side of the graph, terms related to each topic are listed. The blue bars on the right side of the graph show the frequency of the term in the corpus, while the red bar graph that appears when a topic

is clicked represents the frequency of that term in that specific topic.

In this study, topic analyses were conducted with different numbers of topics. As a result of these analyses, it was observed that there were distinguishable and meaningful topics, so the five-topic model was preferred.

In the visualization table of the topic analysis obtained in the scope of the study, the circles of the five headings obtained are shown (Figure 8). These topic circles are drawn using a dimensionality reduction technique called PCA. The aim here is to ensure that each circle has a distance to prevent overlap and to make each circle unique. As the visualization is interactive, when hovering over a circle, different words are displayed on the right side of the table, and the frequency of the selected topic is marked in blue and the estimated term frequency in red. It can be said that topics that are closer to each other consist of more relevant contents (13). As can be seen in Figure 8, the corpus has diverged into five different topics and the obtained expert theses are grouped under five different topic headings.

The distance map between topics, the values related to 30 terms associated with each topic, and the pyLDAvis graphics showing the word frequencies in the corpus and the topics 1, 2, 3, 4, and 5 are shown in Figures 9-13.

Accordingly, in the word-based evaluation conducted in the first topic, it was determined that the top five concepts that are most frequently mentioned in terms of frequency are "score," "pregnant," "scale," "datum," and "pregnancy." Additionally, it was found that almost all of the words "scale," "education," and "anxiety" in this corpus are included in this topic, and the words "form," "mother," "life," "information," "sexual," and "health" are also more commonly used in this topic compared to other topics. In terms of the concepts included in this topic, it was evaluated that the topic indicates the academic aspect of the expert theses, primarily including the forms used, the questionnaires used in the studies, and the literature reviews conducted. Therefore, it was concluded that this topic could be titled "**Academic Research.**"

In the word-based evaluation conducted in the second topic, it was determined that the top five concepts that are most frequently mentioned in terms of frequency are "pregnancy," "pregnant," "fetal," "week," and "preeclampsia." It was also found that almost all of the words "preeclampsia," "preterm," "artery," "doppler," and "neonatal" in the corpus are included in this topic. Furthermore, it is seen that the word frequency of "maternal," "gestational," "cesarean," and "week" is

more dense in the second topic than in other topics. In this context, it was evaluated that this topic could be titled "**Neonatal Studies**" based on the obtained word frequencies.

In the word-based evaluation of the third topic, it was determined that the top five most frequent concepts were "pregnancy", "control", "treatment", "ovarian", and "day"; however, almost all of the words "cycle", "pco", "pcos", "rat", "amh", "follicle", "hormone", "ovary", "oocyte", and "ivf" were included in this topic. It was also determined that the words "infertility", "insulin", and "endometriosis" appeared more frequently in the third topic than in other topics. In this context, it was evaluated that this topic could be titled "**Studies on Reproduction and Ovarian Treatment.**"

In the word-based evaluation of the fourth topic, it was determined that the top five most frequent concepts were "cancer", "case", "cervical", "hpv", and "test"; almost all of the words "hpv", "tumor", "survival", "gene", "invasion", and "node" were included in this topic. However, it was also determined that the words "cancer", "gdm", "expression", and "vitamin" appeared more frequently in the fourth topic than in other topics. In this context, it was evaluated that this topic could be titled "**Cancer Research.**"

In the word-based evaluation of the fifth topic, it was determined that the top five most frequent concepts were "endometrial", "operation", "postoperative", "treatment", and "surgery"; almost all of the words "operation", "postoperative", "hysterectomy", "urinary", and "incontinence" were included in this topic. However, it was also determined that the words "surgery", "pelvic", "surgical", "preoperative", and "bleed" appeared more frequently in the fifth topic than in other topics. In this context, it was evaluated that this topic could be titled "**Hysterectomy and Other Surgical Procedures.**"

As a result of the Topic Analysis of 3,344 specialist theses registered in the National Thesis Center, it was determined that the studies were grouped under the following headings:

1. Academic Studies and Research
2. Studies on Newborns
3. Reproduction and Ovarian Treatment Studies
4. Cancer Research
5. Hysterectomy and Other Surgical Procedures

The study conducted has identified that the specialization theses registered in the Turkey Thesis Center are grouped into five main topics in the context of "**Obstetrics and Gynecology**". It is considered that the evaluation of the future specialization theses in this

context and taking this into consideration in the selection of the topic could contribute more to the literature.

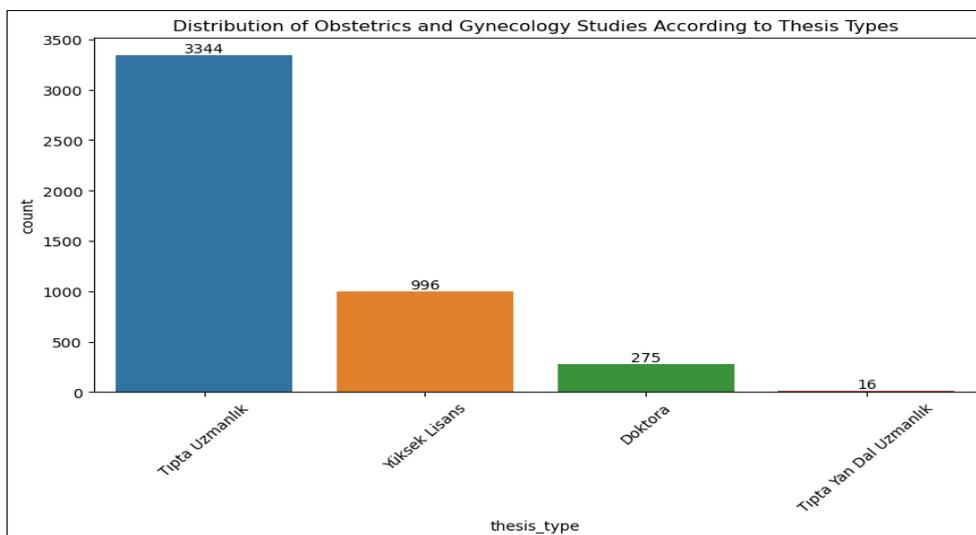


Figure 1: Distribution of Theses by Type

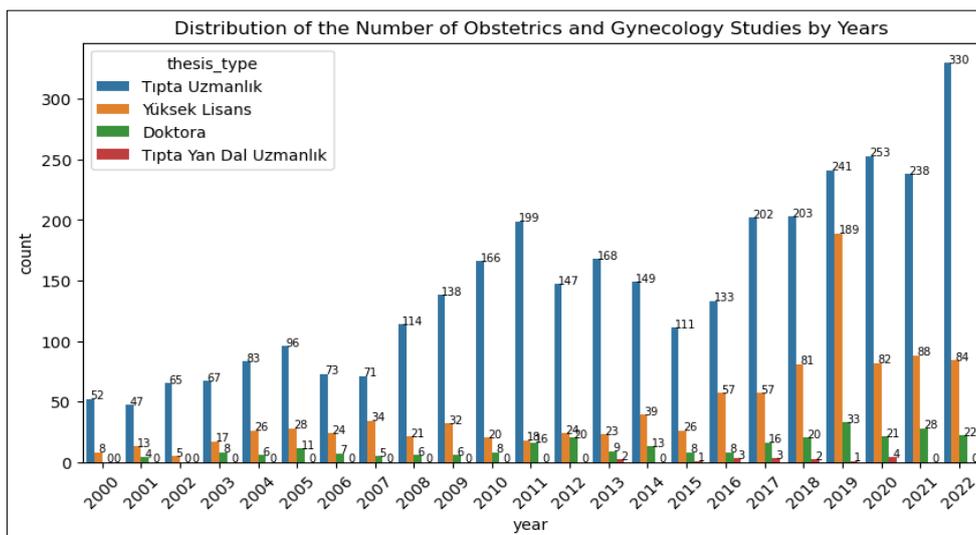


Figure 2: Distribution of Studies by Year

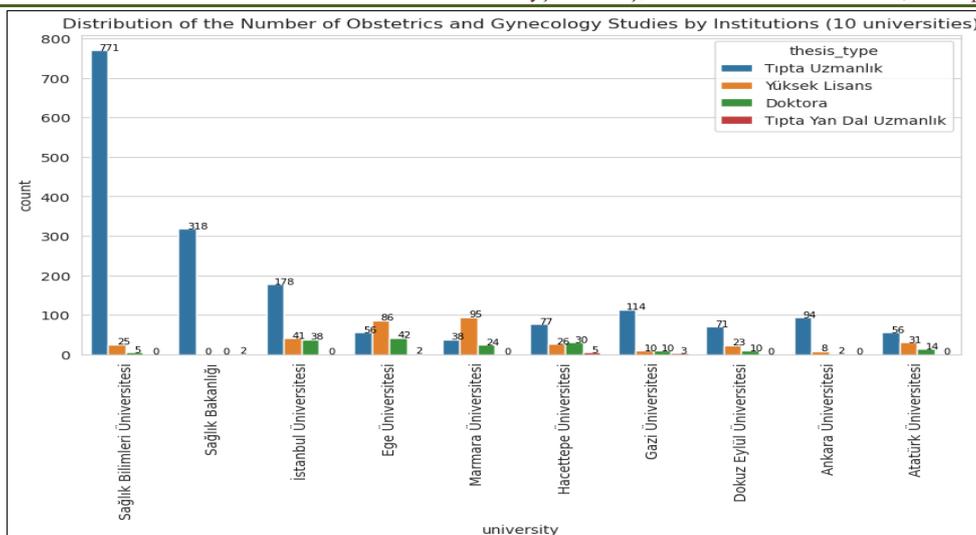


Figure 3: Distribution of Theses by Institution (Top 10 Universities)

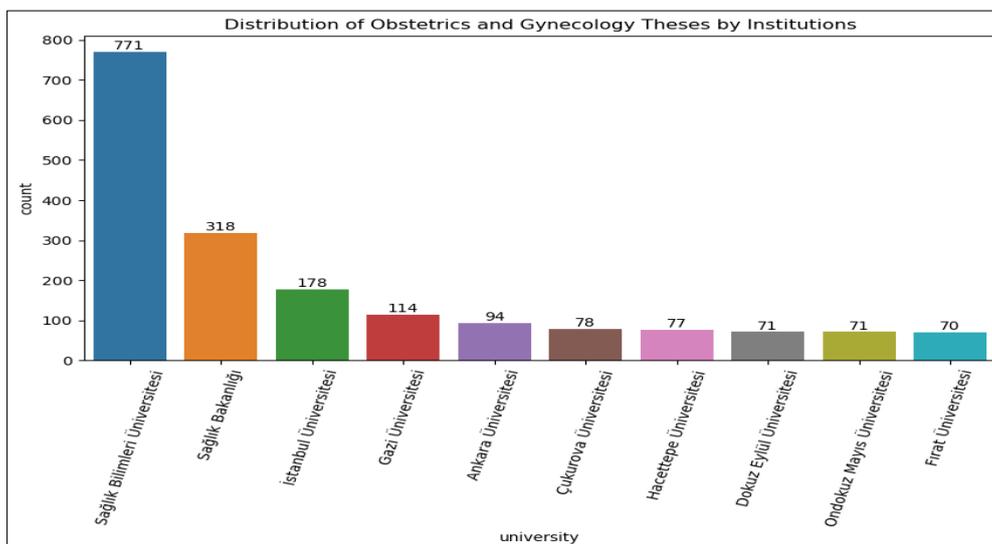
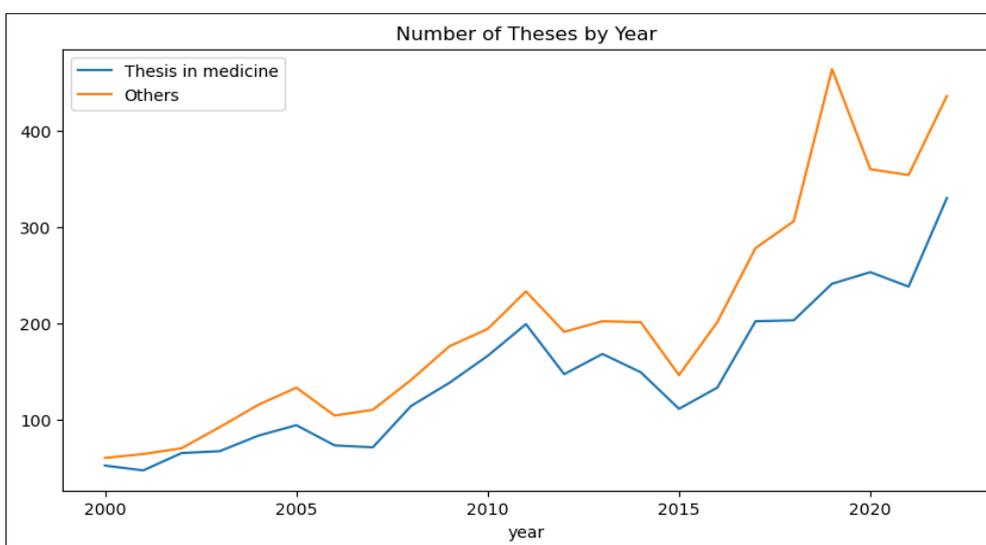


Figure 4: Distribution of Specialization Theses by Institution



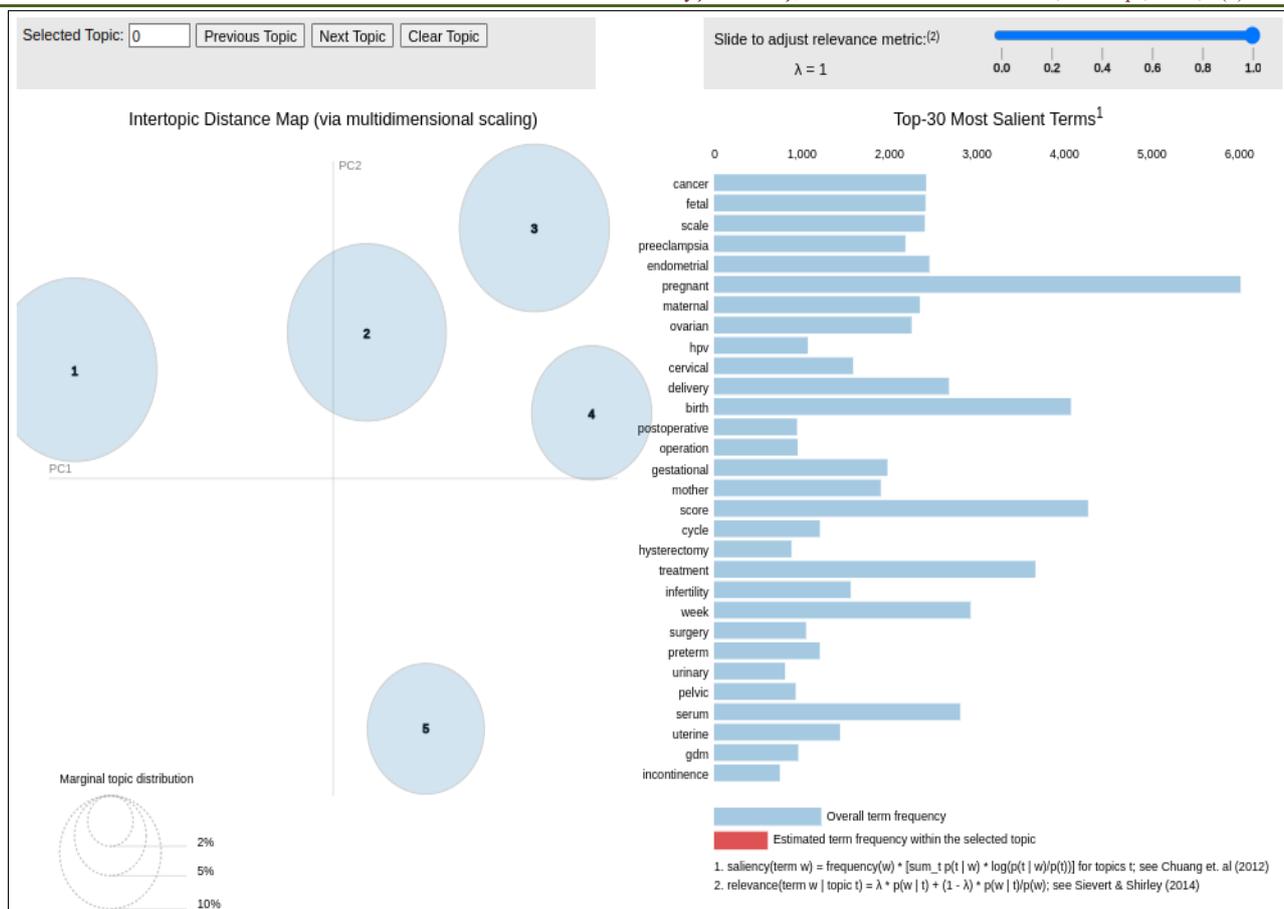


Figure 8: pyLDavis Visualization Graph for the Research Model, with 5 topics

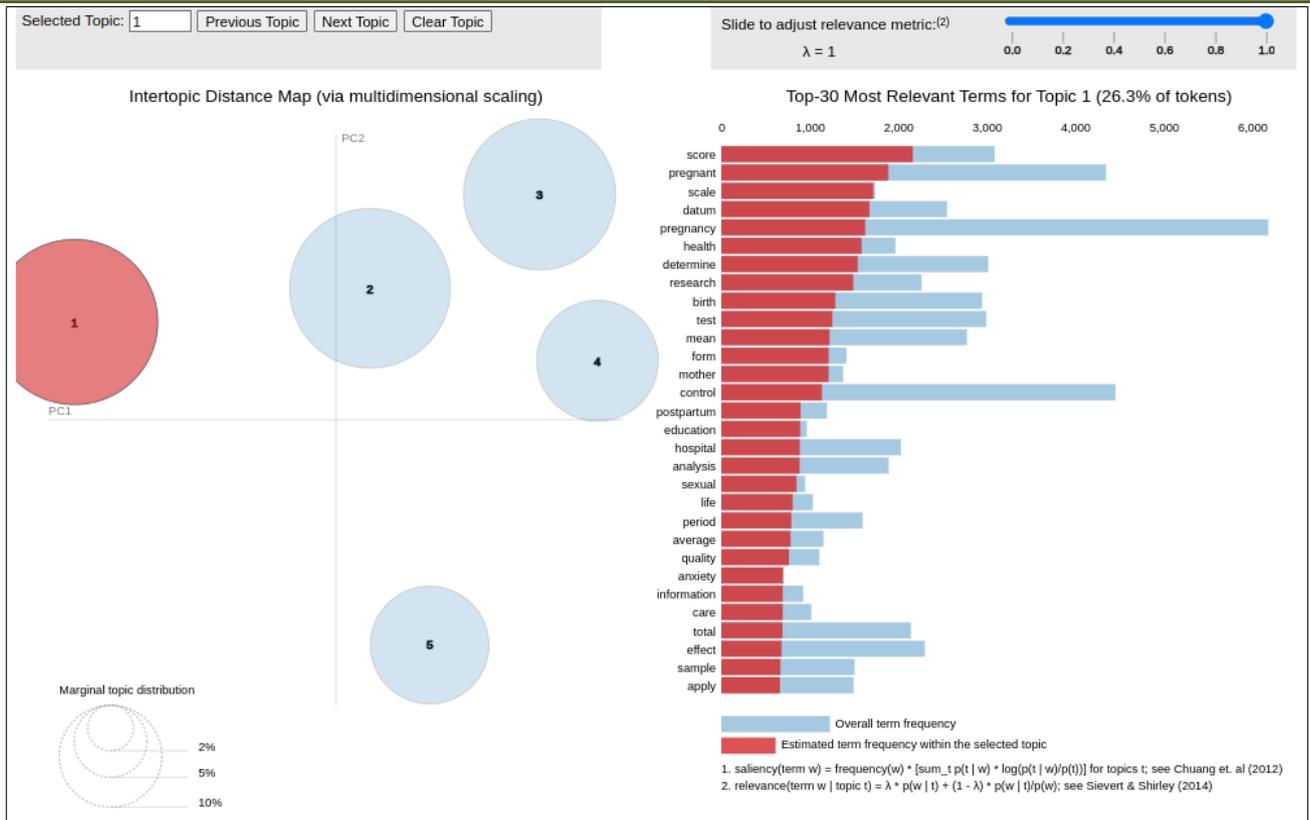


Figure 9: pyLDAvis Visualization Graph of the First Topic

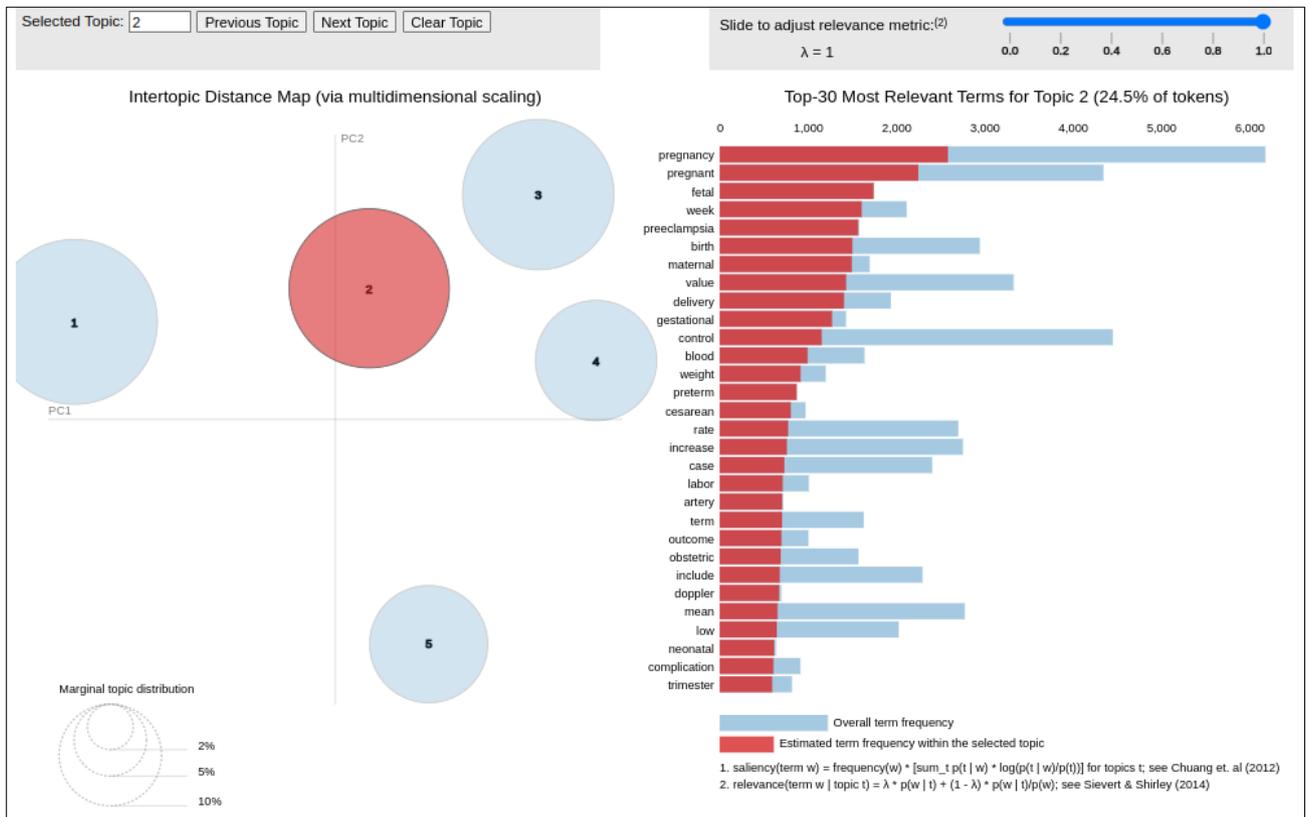


Figure 10: pyLDAvis Visualization Graph of the Second Topic

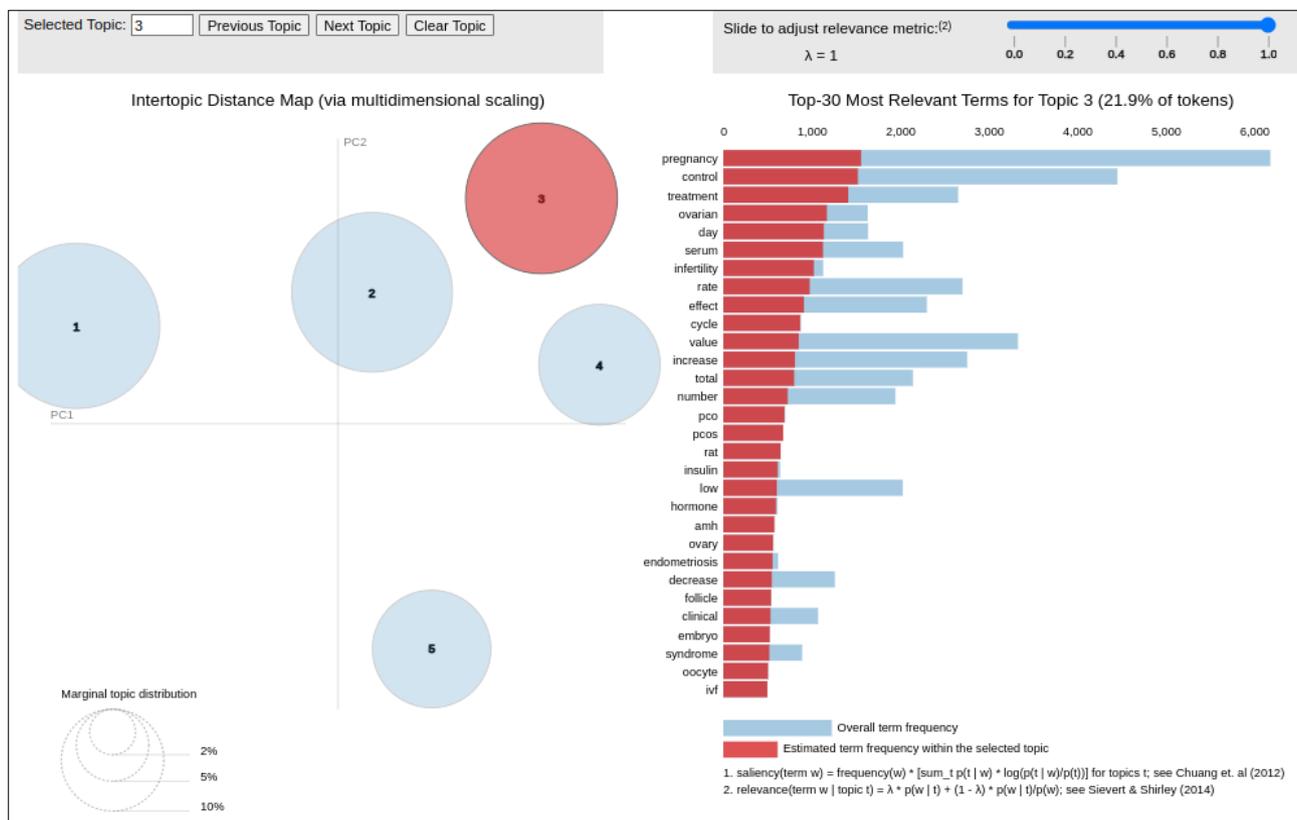


Figure 11: pyLDAvis Visualization Graph of the Third Topic

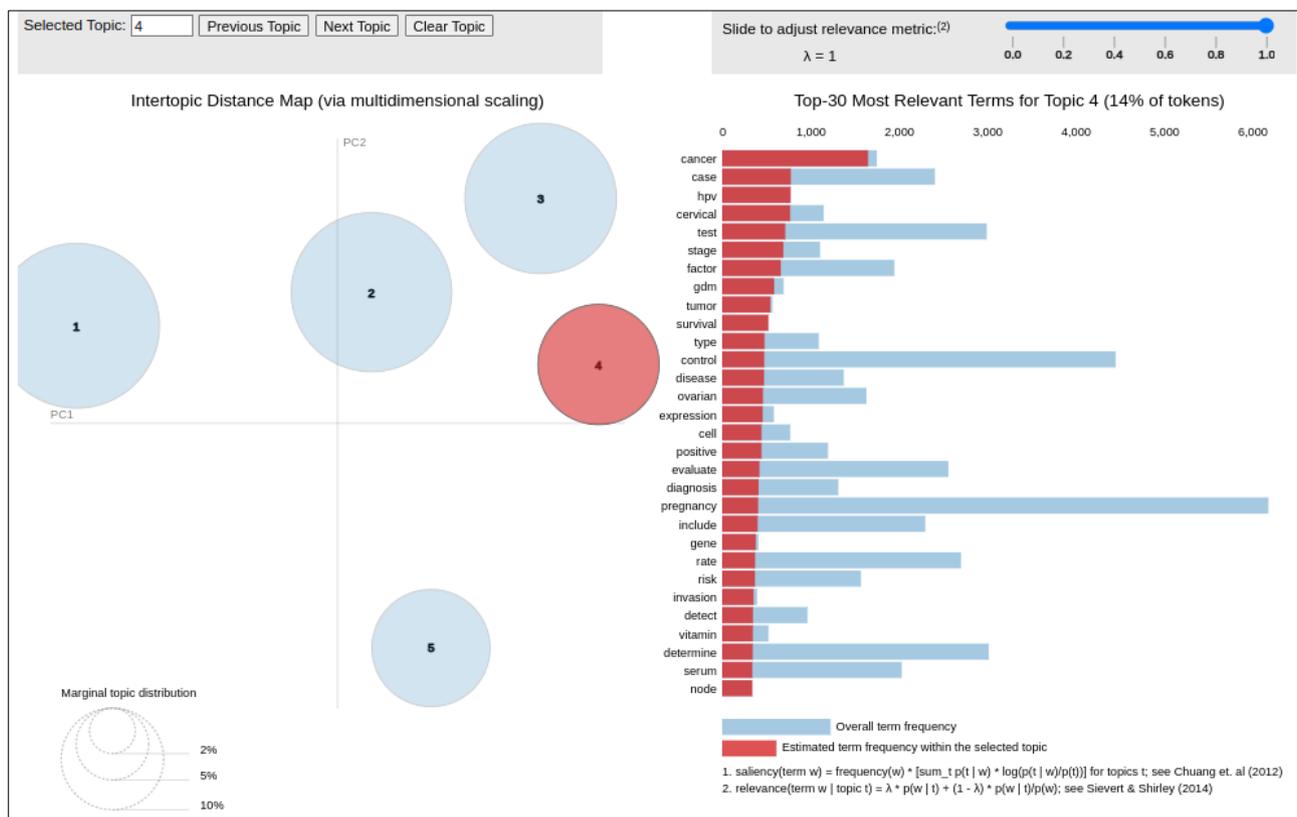


Figure 12: pyLDAvis Visualization Graph of the Fourth Topic

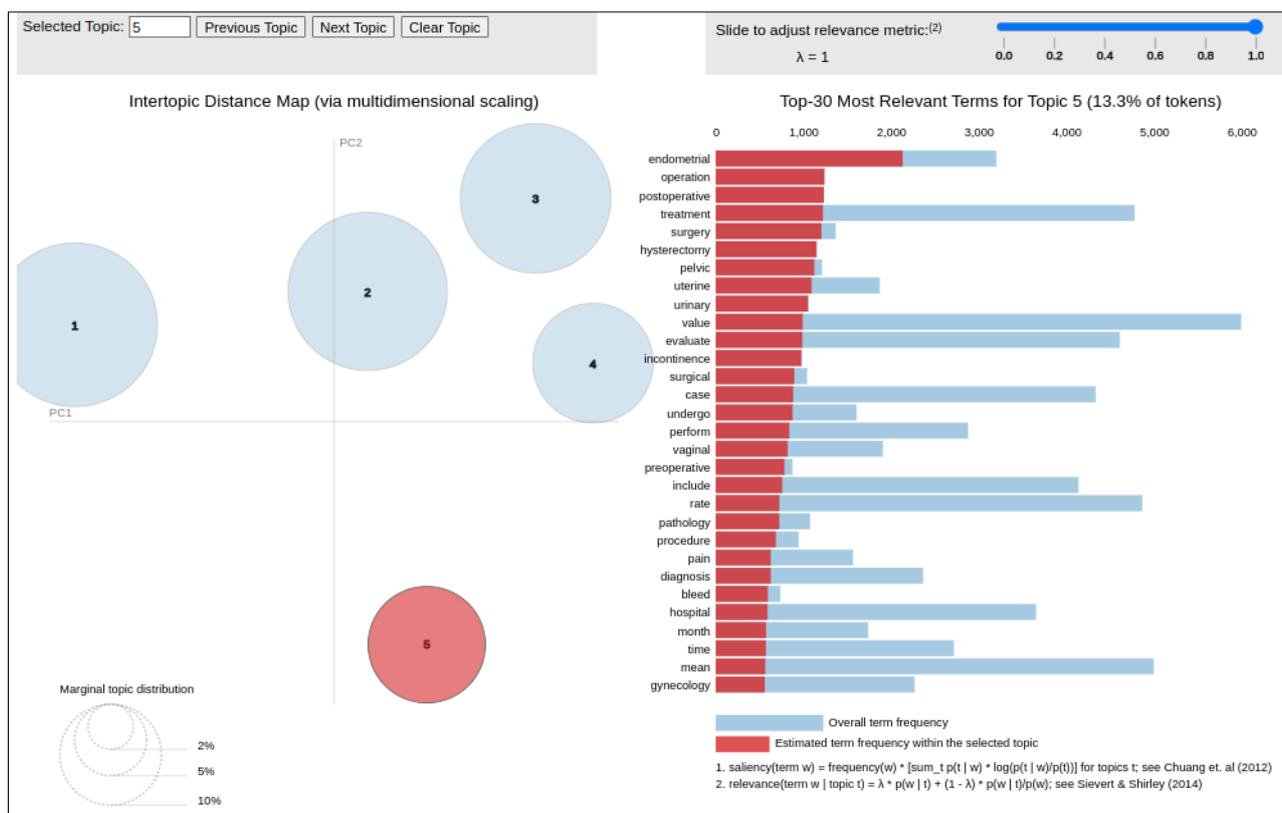


Figure 13: pyLDAvis Visualization Graph of the Fifth Topic

CONCLUSION

The study presented in this paper conducts an in-depth analysis of the thematic landscape encapsulated within specialization theses concerning "Obstetrics and Gynecology" archived in the Thesis Center in Turkey. From the vast repository of 759,771 theses documented in the National Thesis Center since its establishment in 1959, a selective examination was undertaken, focusing solely on the 4631 studies pertaining to "Obstetrics and Gynecology". Employing the Python programming language facilitated the statistical analysis of the collected data, while topic modeling was executed utilizing the Gensim library and the Latent Dirichlet Allocation (LDA) algorithm. The findings elucidate a discernible concentration of research endeavors within the domain of obstetric and gynecological spheres, encompassing aspects such as obstetric and gynecological diseases, infertility, surgical interventions, malignancies, antenatal care, delivery management, perinatology, prenatal diagnostics, and ultrasonography.

This scholarly inquiry extends its scope to evaluate the thematic orientation of theses published in the field of Obstetrics and Gynecology in Turkey, encapsulating the temporal spectrum from 2000 to 2022. The corpus encompasses 3,344 medical specialization

theses, 996 master's theses, 275 doctoral theses, and 16 medical sub-specialization theses. Notably, the zenith of thesis productions was registered in 2019, with the Health Sciences University emerging as the preeminent institution in terms of thesis output. A notable trend is observed wherein until 2017, a partial parallel linear relationship between medical specialization theses and other thesis genres existed, yet since then, a surge in the numerical value of the latter has transpired.

Furthermore, the discourse conducts an in-depth examination of word frequencies and their semantic attributes across five distinct thematic clusters within the corpus. Notably, the prevalent topics encompass academic research, newborn studies, reproductive and ovarian treatment studies, cancer research, and surgical interventions. A comprehensive understanding of the thematic landscape indicates a significant emphasis on multifarious aspects of obstetric and gynecological care and research.

In conclusion, this study furnishes a meticulous scrutiny of the thematic underpinnings characterizing specialization theses in Obstetrics and Gynecology in Turkey. The findings underscore a prevalent focus on key domains within the field, while also illuminating

potential avenues for future research endeavors. Recommendations include diversifying research horizons, incorporating psychological and social dimensions, and exploring the ramifications of novel technologies in women's health. Nevertheless, it is imperative for subsequent investigations to embrace larger sample sizes and employ diverse analytical methodologies to foster a more nuanced understanding of the subject matter. This study stands as a seminal contribution to the scholarly discourse surrounding specialization theses in Obstetrics and Gynecology and serves as a beacon for future research endeavors in this domain.

Study Limitations

It is important to acknowledge certain limitations inherent in this study. Firstly, it should be noted that academic theses in Turkey are digitally archived in the National Thesis Center (UTM) (4) and are made available as open-source materials for academic research. As of January 07, 2023, a query conducted in the UTM revealed a total of 759,771 registered studies since 1959. However, for the purposes of this study focusing on Obstetrics and Gynecology, only 4631 relevant studies were included.

Funding Source: No financial support was received for this study.

Competing Interests: The authors have no conflicts of interest to declare.

REFERENCES

- Leddy MA, Lawrence H, Schulkin J. Obstetrician-gynecologists and women's mental health: findings of the Collaborative Ambulatory Research Network 2005–2009. *Obstetrical & gynecological survey*. 2011 May 1;66(5):316-23.
- Bruno AM, Blue NR. Challenges in interpreting obstetrics and gynecology literature. *Clinical obstetrics and gynecology*. 2022 Jun 1;65(2):225-35.
- Tonta Y, Akbulut M. Türkiye’de lisansüstü tezlere açık erişim. *Türk Kütüphaneciliği*. 2019 Dec 27;33(4):219-48.
- YÖK. (2023). *Ulusal Tez Merkezi*. <https://tez.yok.gov.tr/UlusalTezMerkezi/>. Erişim tarihi: 07.01.2023
- Python. (2022). *Python.org*. (<https://docs.python.org/3/license.html>) . Erişim tarihi: 23.01.2023
- GENSIM. (2023). *GENSIM, topic modelling for humans*. <https://radimrehurek.com/gensim/intro.html#what-is-gensim>. Erişim tarihi: 23.01.2023
- Kapadia S. Evaluate topic models: Latent Dirichlet allocation (LDA). *Towards Data Science*. 2019 Aug 19.
- Kapadia S. Topic modeling in python: Latent dirichlet allocation (lda). *Towardsdatascience.com*. 2019 Apr.
- Aydın G, Hallaç İ. Türkçe Metinlerde Otomatik Konu Tespiti. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*. 2021 Sep 9;33(2):599-606.
- Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of machine Learning research*. 2003;3(Jan):993-1022.
- Chuang J, Manning CD, Heer J. Termite: Visualization techniques for assessing textual topic models. In *Proceedings of the international working conference on advanced visual interfaces 2012* May 21 (pp. 74-77).
- Sievert C, Shirley K. LDavis: A method for visualizing and interpreting topics. In *Proceedings of the workshop on interactive language learning, visualization, and interfaces 2014* Jun (pp. 63-70).
- Ghanoum T. Topic modelling in Python with spaCy and Gensim. *Towards Data Science*. 2021 Dec 20.
- McLevey J. *Doing computational social science: a practical introduction*. Sage; 2021 Dec 15.
- Tran K. pyLDavis: Topic Modelling Exploration Tool That Every NLP Data Scientist Should Know [Internet]. 2023 [cited date, if available]; Neptun.Ai, MLOps Blog. Available from: <https://neptune.ai/blog/pyldavis-topic-modelling-exploration-tool-that-every-nlp-data-scientist-should-know>.