CSSM 530

AUTOMATED TEXT PROCESSING FOR SOCIAL SCIENCES:

Unpacking 'Perceptions': A Deep Dive into the Discourse Surrounding the Turkish Education System

AYSE ZEYNEP ZULFIKAR

Prof. Dr. Fatoş Gökşen

JUNE 2024

1 Abstract

This study embarks on a journey into the public's perceptions of the Turkish education system, a topic that has yet to be fully explored in existing literature. By employing Latent Dirichlet Allocation (LDA) to analyze data from Ekşi Sözlük, a popular Turkish online forum, we aim to delve into society's views on the education system, its perceived issues, and the proposed solutions. This exploration will help us understand the societal connotations of 'education' and the 'education system.' Our project adopts a data-driven and exploratory approach, for trying to minimize bias and allowing the data to guide our understanding. Although we acknowledge potential limitations due to the complexity of the data and the exploratory nature of our study, our main focus is to gain a comprehensive understanding of public perceptions on education-related issues in Türkiye. Furthermore, this paper provides beginner-friendly insights into the practical application of LDA as a tool in Natural Language Processing (NLP), highlighting its capabilities and constraints in analyzing public discourse.

2 Background

Education system changes, with their accompanying decline in quality of education at all levels, is a topic frequently debated in Turkish public discourse. When thought about in the context of the complex and special place that connotations of education has on Turkish society, that it is not understood only as a process for skills development, but also as a beacon of hope for social mobility, betterment of one's life conditions and diversification of life chances, these discussions become more significant for understanding the social fabric of Türkiye. However, even though there have been studies that tried to explain the economic, social and political influences that shaped the changing dynamics of the Turkish education system and the inherent problems attached to it, most of the focus has been given to more concrete problems.

A quick literature review shows that Turkish education studies are mainly interested in comparing the Turkish system to other national education systems (Balcı, 2007; Telci, 2011; İş, 2017; Ömeroğlu et al., 2024) or curriculum issues, test-based instruction and classification of the quality of Turkish education itself through PISA results (Kartal 2015; Çelen et al., 2011; Tertemiz and Aytekin, 2018). These studies also give birth to problematization of management issues within the education system through problems of the need for professional development of teachers in line with EU policies curriculum issues (see Terzi 2014) or educational policy actions by the Ministry of National Education in the times of COVID-19 (see Özer, 2020).

With time, the unique viewpoints of the education system's participants have also started to intrigued researchers, reflecting the need for deepening the analysis for understanding the mechanisms of the system more clearly. For instance, Çelik et al. (2021) recently explored the dissatisfaction and worries of students through sentiment analysis of tweets concerning open and distance education in the aftermath of the COVID-19 related education policies. However, the majority of these studies aiming to comprehend perspectives primarily focus on principals, teachers, and educator candidates to address these 'issues,' identify their 'root causes,' and propose 'solutions. Şener (2018) found student-related issues as primary in Turkish education, followed by finance, curriculum, teachers, and politics. Abu et al. (2016) identified major concerns as teacher appointments, unequal access, lack of supervision, infrastructure, and ambiguities in the new system. Saylık et al. (2021) highlighted perceptions of complexity and suppression of individuality, suggesting stability and liberation. Neyişçi et al. (2020) proposed policy development, societal awareness, program updates, safety measures, and democratic education. Although these studies strive to create a scholarly, hence systematized, interpretation of underlying mechanisms of these debates, public perception of these issues, which fuels the reproduction of these problems themselves, still remains largely unexplored.

3 Introduction

This study seeks to bridge the gap in the Turkish literature concerning public perception of the education system by investigating how Turkish society perceives education systems as well as how they formulate problems within them and propose solutions for it. To accomplish this goal, we plan to explore the interpretations of 'education' and 'an education system' from their viewpoint, utilizing data from Ekşi Sözlük topics shared on the platform under the category of 'education system'. Through employing a popular topic modeling algorithm for NLP, Latent Dirichlet Allocation (LDA), we will identify key themes and patterns in (1) how individuals perceive education system, (2) see which problems in it, and see (3) how they propose solutions these challenges. For this paper, we are driven by a desire to understand the underlying connotations of education for people; our hope is that by discerning what individuals perceive as the 'real problems' within the education system, and by considering their proposed solutions, we can gain insight into what they identify as barriers to their anticipated outcomes from education. Understanding how they formulate solutions provides a glimpse into their ideal scenarios.

While we are aware that there may be numerous potential dependent variables identified in the extensive academic literature, and our approach here is somewhat indirect, our aim is to move beyond preconceived ideas and let the data inform our understanding. This is shaping an exploratory approach for us to navigate through the data. We take this approach through the help of an unsupervised learning technique which will give us 'the opinion' of the public without any prior filter. Despite this goal being ambitious and the complexity of the data suggesting a reading for you that lacks a valid, reliable analysis for robust generalizations about Turkish people's perceptions of education-related issues, we are more focus on underlying the features of the paper that is (1) enabling you to observe an NLP amatour's attempt to understand LDA modeling and use it as a tool to decipher public perception without the bias of former issue formulations by domain experts, which often dictate 'how things are' without leaving space for alternative explanations, and (2) follow up the journey of a sociologist's struggle to try to fill up a gap in social science literature without the initial establishment of variables to direct the course of the analysis or relying on comfortable use of traditional methods to do so.

4 Method and Data

4.1 Data

As this study aims to fill a gap in the literature regarding public perception of the Turkish education system, it's essential to explore interpretations of 'education' and 'an education system' from their perspective, identifying the underlying connotations and connections. Ekşi Sözlük serves as the platform for this purpose, acting as the most popular forum for sharing opinions on diverse topics, including education. The platform operates like any other forum; a topic is introduced, and entries are submitted under it. What sets Ekşi Sözlük apart are the 'channels' that are constantly updated with the latest topic shared on the platform related to that subject, covering general topics that are publicly discussed (like politics, sports, relationships, education); however, we will exclude this feature from our analysis and focus only on gathering all of the topic ever shared on the platform that mentions 'education system.'

The exploratory approach we've designed for our study necessitates a deep dive into a methodology that equipped to mitigate potential bias regarding the anticipated outcomes of the model. Although it's virtually impossible to create conclusions for this study without basing them on our understanding of the issue, we must utilize unsupervised Natural Language Processing (NLP) tools to establish minimum bias in potential topic classifications within the data. While discussions about 'bias' in social sciences can often be perceived as monotonous, largely due to the accepted reality among social scientists that nothing 'said' can be completely free of bias, we, as beginners to computational analysis in social science, are enthusiastic about the potential of our 'unbiased/unfiltered/unsupervised' approach can bring to us. Especially, by applying this approach to the complex question of 'public opinion', which necessitates navigating a complicated web of meanings that most of the time have a tendency to contain contradictory statements, we hope to craft a project that can perhaps shed light on some facets of this intricate web - for this purpose, we want the model to perform this task for us without human intervention, to not limit and reduce the model's understanding of these contradictions to simpler forms.

Since models like BERT, RoBERTa, FastText, and transformer models require a high level of annotation and pre-training, which includes many instances of 'intervention' that could introduce 'bias', we will not consider them as options for this study. Among the most popular unsupervised models, which include Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), Word2Vec, GloVe, Doc2Vec, and Hierarchical Dirichlet Process (HDP), LDA stands out for a beginner's first foray into NLP as it requires the least intervention on the model itself after data cleaning and preprocessing, compared to rest. Since LDA is a form of topic modeling, to distinguish between the topics that the modeling will identify in our data and the topics for which Ekşi Sözlük has entries (and creates the dataset we will use), it is possible to get confused. From now on, when we refer to topics, we talk about what the model will find after we work through entries. The paper is, from now on, about finding the repeated topics within these entries.

What LDA will do here is, as a commonly used topic modeling algorithm in NLP, it will identify hidden topics within our collection of entries we took from Ekşi Sözlük using a probabilistic model. As an unsupervised learning technique, it will reveal the topics found within each entry, each represented by word distributions, and the model will display these word distributions for each topic it identifies. The model's strength comes from its reliance on the Dirichlet distribution, which is again a probability distribution used to model the distribution of topics in each entry and the distribution of words in each topic. The LDA process begins by assuming the necessary conditions for this distribution to function. Each entry is considered a mixture of hidden topics from 'casual' conversations people have about education, and these entries also contain word distributions of these topics. For each entry, the LDA model first selects the distribution of topics, randomly chosen from a Dirichlet distribution with parameter α. Afterwards, LDA model will do two things for the word distribution (parameter β): (1) it will select a topic from the distribution of topics it identified, likely the one that dominates the entry, and (2) it will choose a word for this topic. It will repeat this for each entry until all entries are done. The alpha parameter refers to the entry-topic relationship—for example, a high alpha means that no single topic is specifically prominent and the entry contains every topic in a similar percentage. The beta parameter refers to the topic-word distribution, again a high beta meaning no specific words are representative of the topics at hand in that entry. However, there is an important detail about the model that, spoiler alert!, will be important for us to interpret our findings: LDA assumes each topic here is generated independently of the others. This means one topic identified in one entry used can be the same across all entries.

4.2 Model

4.2.1 Pre-processing

Since the model we will be using is an unsupervised one, which means model will do whole the work for us, we first have to help it out with cleaning the data and get on with pre-processing steps.

1. The first step we took after examining the data was to evaluate the number of topics (that contain the entries in Ekşi Sözlük) and the number of entries each topic has. Adhering to the principle that 'the quality of the output is determined by the quality of the input' which is crucial to remember in computational analyses, we set a threshold depending on the context and excluded topics that had fewer than 10 entries per topic. This was done to prevent the analyses from dealing with outliers or overly sophisticated discussion scenarios.

Initially, we started with 157 topics in the dataset, but after this cleaning process, the count was reduced to 66 topics. Despite this reduction, we did not lose a significant number of entries as the total went down to 11,201 from 11,458.

2. At the onset of the preprocessing details, it's necessary to tokenize the text into words so that the Python function can understand the units it will work on. It's crucial to perform tokenization before any other steps, as the code may not function without it, or even if it does, it may not process the entire text as required.

3. Upon reviewing the data, we performed a general clean-up, removing punctuation, numbers, and single characters. We also identified newlines, extra spaces, and additional single characters. To eliminate these, we utilized the regex function in Python.

4. We lowercase the text as well to get ready for the next step.

5. We eliminated the stop words, utilizing both the nltk.stopwords function and the trstop list from Github.

6. The decision between stemming and lemmatization was a critical factor for the next step. While stemming was relatively fast when implemented, it resulted in meaningless common base roots because it merely removes the last few characters (suffixes) of a word to leave only the stem. If done correctly, this can simplify the analysis. However, in this case, stemming resulted in the repetition of many words in the model output due to the same stem being used in different ways. The stemming model that was used for the Turkish language (turkish-stemmer-python from Github) did not perform efficiently. Due to these reasons, stemming was not included in the pipeline of this project.

7. As an alternative, we decided to do lemmatization using the Zeyrek package from Python, which is a morphological analyzer and lemmatizer for Turkish. It proved to be more accurate than stemming in terms of identifying meaningful dictionary words, as it takes context into consideration. The practical difference between stemming and lemmatization is that while stemming simply removes common suffixes from the end of word tokens, lemmatization ensures that the output word is a normalized existing form of the word (for instance, a lemma) that can be found in the dictionary.

8. Subsequently, after examining the word frequencies within the dataset, we made the decision to eliminate words that appeared more than 2000 times. These words were recurrent and did significantly undermine our purpose of revealing the hidden topics we were attempting to uncover from the dataset.

9. As the final preprocessing, as directed by our research question, we divided the 66 topics (those which have entries in Ekşi Sözlük) into three categories pertaining to (1) perception of an 'education system', (2) the issues identified within the system, and (3) the proposed solutions to these issues. This categorization was achieved through the labeling of topics, which were then included as columns in the dataset. Then, the main dataset again was filtered and divided into three distinctive datasets according to each new dataset only containing their respective entries: $\text{system}_{\text{df}}$, $\text{problem}_{\text{df}}$, $\text{solution}_{\text{df}}$.

4.2.2 Model Training

The real work begins with training the model. After completing the preprocessing phase of model construction, the next task has to be optimizing the number of topics to analyze. For this, we experimented with various numbers, resulting in multiple models, some of which made the final cut for the project's final code file. However, using more advanced techniques like GridSearch to decide on the number of topics and words to find within the dataset proved challenging for a beginner's project (promise, we tried!). Thus, we decided to formulate the model in a manner similar to how GridSearch would operate, but without directly using the function. We tested a range of topics from 5 to 25 in this new model, including 1 to 20 passes in this formulation. The model found the optimal version for all three datasets based on coherence and perplexity scores, which will be discussed in detail in the evaluation part of this section.

However, the primary challenge in this project was the analyst's lack of experience in the realm of NLP to effectively train the models. For instance, a sophisticated hyperparameter tuning or successful implementation of bi-grams or n-grams could have led to significantly more reliable model findings; for example, adding bi-grams were tried for this project, but they failed to provide improvements to the models' output due to the unsuccessful implementation. Although TF-IDF was implemented, the lack of advanced coding skills for the model meant that the model's results remained some-what unchanged. Indeed, the main issue with the improvements made was that they could not be finely tuned according to the unique attributes of the dataset and the final requirements, due to the analyst's lack of experience in both NLP models and necessary advanced coding skills.

Nevertheless, this lack of experience meant that we could try to get creative and learn more about the model or the technique we were using, more than we would have if we were more educated on what I was supposed to do. Indeed, the mistakes made can lead to a more comprehensive understanding of how the model operates.For instance, as mentioned earlier, LDA assumes that topics are independent of each other and of entries - a topic identified in one entry can be the same across all entries. However, when dealing with data that continuously repeats the same words across topics, despite preprocessing, it may indicate that either the data is faulty or the training may not have been done correctly. If this problem persists even after eliminating high-frequency words from the dataset, as we did, it necessitates a reconsideration of the current approach. Our solution was to add a line of code that ensures a word cannot be included in more than, for example, three topics. This slightly improved the scores, however, we quickly realized that this approach contradicts the model's main assumptions.

4.2.3 Model Evaluation

The performance of the LDA model is generally assessed using metrics like the coherence score, which measures the degree of semantic similarity between high-scoring words in the topic, and perplexity. The formal definition of perplexity is 'how well a probability distribution predicts a sample,' but it essentially provides a score indicating how well the model can predict the subsequent word in the mechanism - how accurately it can predict the unseen, next word of the topic with the words it has. Even though alpha and beta parameters are also significant details for observing the ratio of entry-topic and topic-word distributions, we did not include them in this model evaluation. This was because we were unable to train the data based on the fine-tuning of hyperparameters due to a lack of necessary understanding of their relationship.

Even though the alpha and beta parameters could have been reported, since they were not fine-tuned, they will not be included in this report. However, the primary reason we are not delving deeply into this model's evaluation is that we couldn't achieve a satisfactory score to analyze the model's performance in detail. For all three models, the coherence scores were remarkably low, with no score higher than 0.38. In the case of the perplexity score, it was surprisingly negative for all three models. However, this does not necessarily suggest a low value. We used Gensim's 'logperplexity' function to calculate the perplexity score, which often results in negative values due to the automatic conversion of small probabilities to the log scale. This is primarily because Gensim does not provide a perplexity score but the likelihood bound that is used in the equation for perplexity's lower limit. Although a lower perplexity score is generally preferred, the lower bound value denotes degradation, causing the score to become negative. This degradation worsens with an increase in the number of topics, as a larger number of topics reduces the perplexity even further. In this context, since the values should always be closer to zero, we can deduce that all our perplexity scores, which ranged between -9 and -11, indicated that the models did not perform well. Perplexity, however, is challenging to interpret by definition. Unlike metrics such as BLEU or BERTScore, perplexity only provides a measure of 'confidence' rather than a measure of semantic or textual similarity. Although we would prefer a model with lower perplexity, it does not offer much insight into the semantic quality of the topics in our case.

Finally, the model's outputs also incorporate weight scores; each word in a topic's distribution is associated with a weight that signifies its probability of occurrence within that specific topic – we can interpret this as a measure of the influence that word has in defining what that topic represents in the broader context of the model. Throughout our process of topic modeling, these weights never surpassed 10 percent, suggesting that, akin to the other performance scores of the model, these weights also indicate that we were unsuccessful in developing a reliable model.

4.2.4 Model Validation: Limitations

In the validation process, we are confronted with a fundamental fact of this project: the models we developed for three distinct questions were simply unsuccessful due to a lack of sophisticated code. Despite using all necessary and available packages for Turkish language preprocessing, working with a local language necessitated a deep understanding of what the model required - a challenge we were unable to successfully meet. Nonetheless, as this failure provides valuable learning opportunities for beginners, we drew inspiration from Zelner and his colleagues' (2022) argument that research and writing should occur simultaneously and are essentially a single process – for us, this integrated process also allowed for the completion of this paper, which is a necessary start for a beginner's journey to NLP to learn about what they did right or wrong. Indeed, given the complexity of computational data analysis for social science questions, it becomes unreasonable to expect immediate and straightforward answers from a beginner – however, documenting the research process through writing, which also facilitates understanding of the thought process behind the project, is crucial for the continuation of the path which will lead to the project's actualization. This understanding provides context and highlights the necessary improvements for the current research and future iterations.

The thought process in this context reveals that the incorrect decisions made during the project's development form the basis of why validating the model will be challenging – the limitations for this project are notably extensive:

1. The first and major error in the project was selecting Latent Dirichlet Allocation (LDA) as the suitable tool for this data. Although we required an unsupervised model for an unbiased interpretation of discussions about the education system (which offered insights into the system, its issues, and possible solutions), the fundamental structure of LDA, with its document-topic and topic-word distribution, is not suitable for the entries of Ekşi Sözlük. It was only after examining the entry-topic and topic-word distributions generated by the model that we realized most entries in Ekşi Sözlük are typically created with a single topic in mind, unlike a normal document.

2. Despite the majority of the dataset being in Turkish, we failed to recognize during pre-processing that many English words were used throughout the entries. This oversight only became evident after finalizing the model, and our attempts to completely eliminate these words were unsuccessful. While multi-language data is challenging to work with, improving the model's ability to handle it is a critical aspect for the next version of the project.

3. In certain instances, lemmatization was ineffective – while we choose against stemming due to the resulting words seeming meaningless without their suffixes, some words after the lemmatization process resembled stemmed words.

4. While the idea of an unbiased model is crucial for bridging the gap in existing literature, it was essential to choose a focus: either enhancing our understanding of unsupervised models with advanced coding skills or using a supervised technique with annotated data and domain expertise. Unfortunately, our choice was misguided. BERTopic would have been a superior choice for this project, considering our current level of NLP model expertise and coding skills.

5. Our data-driven approach, which was implemented without a targeted focus and domain-specific adjustments, led to a less reliable model. It was tasked with addressing a wide-ranging question that demanded too detailed responses. Concentrating on one aspect of the questions we posed (system, problems, solutions...) would have been more conducive to generating an effective answer.

6. Due to the model's lack of sophistication and the aforementioned challenges, these models cannot be validated as they stand. They need to be revised to meet the requirements of a more dependable model, considering the alternative options that could have been selected earlier. However, the output of these models serves as a useful starting point if we aim to fill the gap in the literature. Thus, we will proceed with the results in the next section.

5 Results

As previously discussed, we developed three distinct LDA models for this project. Although somewhat similar, they have differences in word cleaning and the respective number of topics. These variations stem from the data itself, as we ensured that our specific filtering for systemdf, problemdf, and solutiondf produced the right number of topics for analysis.

Model (a) focused solely on topics that specifically addressed the system itself; in this model, we found the highest coherence and perplexity scores were with 20 topics, each containing 10 words, over 15 passes. The coherence score is 0.3, and the perplexity score is -9.4. The distribution of topics and their respective proportions are as follows:

1. Absence, good, to do, only, plain, source, to be, state, job, money
2. Flat, religion, to do, hand, house, to simplify, absence, first, only, plain
3. System, grade, thing, success, found, basic, secular, English, result, math
4. With, to do, clock, it is a system, one, example, money, system, test, result
5. Right, first, result, to do, generation, good, age, method, Turk, big
6. To be, only, plain, absence, to do, big, dung, one, good, money
7. Absence, one, good, special, to be, to do, real, on, education, system
8. Absence, reform, system, good, for, Turk, to be, there is not, shortage, Turk
9. Plain, generation, it is a system, only, right, memorize, to be, single, young, on
10. Big, only, plain, to do, time, result, to be, place, example, value
11. Quality, total, absence, job, unit, month, national, mill, national, mile
12. To do, absence, job, to be, right, teacher, what, world, time, to understand
13. Good, absence, target, to be, system, to do, far, last, beautiful, place
14. Special, science, correct, preacher, average, basic, wrong, test, Imam, state
15. To do, to be, good, graduate, big, first, state, past tense of to say, middle, necessary

16. Good, department, to read, absence, time, job, construction, plain, same, only
17. Big, middle, piece, to do, single, state, sun, picture, to be, time
18. To do, to be, absence, good, first, place, profession, man, job, to read
19. To do, to be, good, time, last, job, absence, special, world, topic
20. History, absence, world, curiosity, to do, day, other, available, to be, to arrive

For model (b), which solely focused on topics related to the problems of the Turkish education system, we achieved the best coherence and perplexity scores with 20 topics, each comprising 10 words, over 15 passes. The coherence score for this model is 0.4, while the perplexity score is -10.6. The distribution of topics and their respective proportions are as follows:

1. You, mountain, license, are, ten, search, new, strange, name, emotion
2. High, state, special, continue, remedy, time, hand, will be, read, money
3. On, established, over, its over, dry, answer, memorize, board, single, in Turkey
4. One, take, week, know, teacher, Anatolia, generation, go, first, understand
5. Politics, missing, history, right, single, life, French, dude, one, one of
6. Same, real, field, life, construction, understand, middle, time, read, learn
7. Teacher, bam, picture, is a system, official, foreign, practical, teacher's, target, system
8. Much, for, not, thing, same, job, student, being, such, its
9. Demand, continuous, target, system, place, supply, individual, appreciation, for, enough
10. Fish, can be, go, technique, incoming, continuous, teachers, ba, tree, bird
11. Mother, pass, turk, world, understand, what, teacher, homework, father, invention
12. Center, place, society, time, education, state, to ask, subject, ball, system
13. Profession, plane, smart, discipline, expert, grown, result, until it ends, method, useful

14. Religion, morality, wane, main, not being, language, history, based, moment, of the system

15. System, not being, shit, is a system, thing, to the thing, continuous, of the thing, la, at the beginning

16. Pattern, graduate, science, young, correct, in, proper, another, memorize, law

17. English, Englishman, English, music, mathematics, Turkish, body, language, Turk, be

18. Society, what, far, aim, construction, state, value, right, owner, free

19. Preacher, imam, imply, special, science, Anatolia, iron, middle, continuous, basic

20. National, mill, flour, national, generation, quality, minister, remedy, continue, administration

For model (c), which solely focused on topics related to the problems of the Turkish education system, we achieved the best coherence and perplexity scores with 25 topics, each comprising 10 words, over 15 passes. The coherence score for this model is 0.3, while the perplexity score is -9.6. The distribution of topics and their respective proportions are as follows:

1. special, real, to be, both, same, middle, world, graduate, good, is a system

2. none, to do, sun, one, to be, understand, profession, does, right, continue

3. none, good, to do, single, system, big, goal, mathematics, hand, very

4. to do, good, new, construction, mother, middle, none, government, primary school, system

5. to do, to be, big, good, job, money, subject, place, over, take

6. none, to do, special, Anatolia, end, new, teacher, to be, first, crap

7. hour, Turkish, simple, only, none, religion, time, government, real, value

8. good, to be, exit, simple, only, system, constant, correct, door, entrance

9. world, subject, good, simple, high, only, learn, to be, to do, judge

10. curiosity, man, to do, pity, other, none, place, invent, throw, island
11. generation, to be, to do, time, guide, good, day, single, missing, logic
12. quality, to be, first, source, to do, good, total, none, place, subject
13. job, good, to be, none, to do, high, field, real, time, big
14. to do, board, to be, time, politics, world, certain, system, owner, education
15. reform, right, government, good, special, other, simple, only, read, place
16. quality, to be, first, source, to do, good, total, none, place, subject
17. hour, Turkish, simple, only, none, religion, time, government, real, value
18. none, to be, to do, place, generation, money, only, simple, right, difference
19. job, to be, to do, big, small, none, value, good, time, coming
20. good, none, money, job, is a system, throw, same, subject, again, can be
21. good, none, money, job, is a system, throw, same, subject, again, can be
22. system, none, young, to do, only, simple, to be, each, end, world
23. job, to be, to do, big, small, none, value, good, time, coming
24. one, history, get, world, same, generation, is a system, to be, mother, only
25. quality, to be, first, source, to do, good, total, none, place, subject

6 Discussion

As previously discussed, there is a gap in the literature regarding our focus on the discussions surrounding the Turkish education system. The closest related work was conducted by Can (2015), who performed a qualitative analysis addressing our key inquiries. However, his study didn't emphasize public perception, and instead concentrated on the viewpoints of administrators and teachers. Consequently, we lack a considerable amount of comparative literature for our findings. Upon revisiting our dataset, we found that even prior to modeling and during the initial analysis of the most frequently appearing words prior to pre-processing, the data provided a synopsis of our intended research focus. The three most frequently used words across all entries were "eğitim", "sistemi", and "yok". In both Turkish and English, these words form a complete sentence, "Eğitim sistemi, yok" (There is no education system). The constant change in the education system, coupled with inadequate policy amendments, has led to a general distrust towards the system as people are unable to perceive a consistent structure. This finding provides a crucial starting point for understanding the rest of the dataset with our current model.

The outputs from model (a) will provide insights into people's perceptions and discussions about the education system. We identified diverse yet complementary perspectives. The actions of 'to do' and 'to be' frequently emerge across the topics, summarizing what education signifies for the Turkish population. While 'to do' underscores skill development for job performance, 'to be' lends support to our argument about social mobility, first introduced in the Introduction section. There is an ongoing pursuit of 'becoming' through education, a means for individuals to achieve their desired identities. In certain contexts, this concept of 'becoming' was interpreted as 'to arrive.'

For model (b), the output, which reveals the perceived issues within the Turkish education system, is a bit more complex compared to the previous one. Despite being given more specific entry groups compared to the first model, the output is somewhat more convoluted. For example, topic (b3) directly examines the issue of education being overly reliant on memorization. Topic (b5) revisits the theme of politics, absence of things, history and rights found to be connected. Topic (b6) explores the concept of truth within the context of 'life' and 'time', with activities such as 'understanding', 'reading', and 'learning' considered within this framework.Topic (b7) discusses art education, with a particular focus on painting classes and the role of teachers. Topic (b9) highlights the mechanism of supply and demand, with individuality and appreciation being identified as linked concepts. Topic (b12) delves into the role of parents in education, particularly in connection with the concept of 'teacher', and appears to be related to homework. Topic (b14) uncovers the relationship between religion and morality and the importance of language and history is underscored, with these elements being foundational.

For Model (c), which focuses on understanding proposed solutions to identified problems, the dominant terms are 'rights,' 'absent,' 'to be,' and 'to understand.' Unlike the previous two models, Model (c) lacks a distinct answer scale, making it less meaningful in comparison. Overall, the frequent use of words like 'quality,' 'good,' 'right,' and 'value' across all topics highlights a desire for high-quality education and concerns about the current system's standards. The repetition of terms like 'state (government),' 'reform,' and 'politics' underlines the acknowledgment of the government's role in shaping the education system and a need for structural reforms. The responsible party for perceived solutions, or the one who will change things, seems to be the government itself, reflecting the paternal figure given to the government ('devlet baba' for a more general interpretation).

7 Conclusion

In conclusion, the limitations of our LDA model, particularly its unsuitability for the single-topic nature of Ekşi Sözlük entries, hindered the validation process, revealing that our findings might not be reliable for broad generalizations. However, despite their flaws, the model outputs offer a valuable starting point for understanding public discourse on the Turkish education system.

Model (a) aimed to uncover what 'education system' meant for Turkish people, revealing a strong desire for education as a means to achieve success and social mobility. This model highlighted concerns about memorization-based learning, the quality of education, and the influence of the government's mindset on the system. Model (b) delved into specific problems perceived within the Turkish education system, again emphasizing concerns about memorization and the influence of politics, while also introducing the roles of art, individuality, parental involvement, and the connection between religion and morality in education.

Mirroring the first two models, Model (c) sought to understand proposed solutions to these problems. Across all three models, a consistent theme emerged: a desire for a high-quality education system and the perception of the government ('devlet baba') as the responsible party for enacting necessary changes. This sentiment is evident in the frequent appearance of terms like 'quality,' 'good,' 'right,' 'state (government),' 'reform,' and 'politics.' The analysis, despite its limitations, sheds light on the perceived shortcomings of the education system, particularly its emphasis on rote learning and the perceived influence of political agendas. The findings highlight a gap between public expectations and the current state of the education system, emphasizing the need for structural reforms and a renewed focus on quality and accessibility.

Even though we were able to see certain aspects of the perception of education system with its problems and provided solutions, we were not successful in revealing the hidden layers of this discussion as we hoped we would. For our highly ambitious aim to be actualized, we need to come up with alternatives to this model or find complementary methods to delve deeper into the discourse. Even though our model provided a decent starting point, it fell short in capturing the nuanced undertones and the multifaceted nature of public opinion. To actualize our ambitious goal, we might need to consider alternative models that are more suited to unearthing the hidden layers of such complex discussions.

8 For Further Research

Each project, with its distinct queries, necessitates a unique perspective on methods and strategies to form the overarching approach - one that is custom-fit to its specific objectives and the characteristics of the dataset upon which all is constructed. However, beginners often face substantial limitations due to an insufficient understanding of the scope this perspective should encompass to form the final structure of the research design. In our case, even our pre-processing steps, which could've been executed in various ways, were missequenced due to unfamiliarity with the method and the dataset. Upon reviewing the final results, for instance, we encountered issues such as insufficient lemmatization (some words were either not meaningful or remained unchanged from their pre-lemmatization format), and the persistence of high-frequency words in the dataset that we had previously removed. Therefore, to address these current limitations, we need to reassess the sequence of steps within the pipeline we formulated for the model, before we start refining the model itself. While it's true that using an LDA model for Ekşi Sözlük entries is something we seriously need to reconsider, upon reflection, we see that creating three separate models to address the individual project questions proved beneficial; despite the challenge of entry-topic distribution, which hindered our initial efforts to fine-tune the code according to alpha and beta hyperparameters, we managed to obtain some form of topic distribution for our three distinct questions - the system in general, problems in the Turkish education system, and potential solutions to these issues. While we persist in questioning the generalizability of these results, we can still discuss the pre-processing and the pipeline of the over-all project to see what could've been done to improve the models.

In a beginner's NLP project, it's essential for the learner to understand the difference between the fundamental aspects of model building and its further sophistications. While data cleaning and preprocessing actions are clear prerequisites for the model to function, fine-tuning the model's code represents the more advanced aspects. These fine-tunings or improvements in the benchmarks requires an detailed eye that cares for the ouput, just like an artist does for its art. However, even though we do have the possion of an artist in this case, we lack the necessary skills for such sophistications – then the focus should be re-directed to what we know about - we need to ensure that the preprocessing is conducted in the correct sequence to accommodate the unique attributes of the dataset. In this context, we can start by examining tokenization and lemmatization as our main problem starts with that. After tokenization and getting rid of regular expressions, next step have to be lemmatization. However, it's crucial to remember that lemmatization depends heavily on correctly identifying the part of speech for each word, which might not always be accurate, especially for languages with complex morphologies like Turkish. Thus, it is worth considering the use of a morphological analyzer designed specifically for Turkish language to improve the lemmatization step in your preprocessing pipeline; despite having utilized Zeyrek, which includes a morphological analyzer to aid in the lemmatization process, we still encountered considerable issues with the output. Hence, suggesting alternatives to Zeyrek or employing a combination of techniques could be vital in this context. Next, eliminating stopwords, commonly used words, and domain-related words is crucial, however, we also encountered with problems in this process, which should've been guided by a thorough understanding of the data and dataset. Using a tailored list of words specific to our study, in addition to the commonly used Turkish stop word lists, is important.

Once the preprocessing is complete, attention must then be shifted towards building and fine-tuning the model itself. As we know, model construction starts with picking the appropriate model that aligns with the data characteristics and research objectives. In our case, although we concluded that LDA wasn't suitable for our data, since we still aim to use an unsupervised model for unbiased analysis, we have to come up with an alternate solution. Initially applying an unsupervised learning model to gain a general understanding of the discussions, followed by supervised methods to delve to the conversation deeper, could be an alternative approach. We could also follow the cross-validation path; while cross-validation is an already employed strategy to bolster a model's robustness, employing hybrid methods to answer a social question or using multiple methods and metrics to discuss robustness is becoming increasingly popular and relevant in today's data age. Nevertheless, fine-tuning the LDA model is still critical to handle these ambitious plans. In this context, various aspects, especially the selection of suitable hyperparameters, need to be reassessed for model sophistication. Focusing on hyperparameters like the number of topics are crucial, and techniques like GridSearch or random search can help determine the optimal hyperparameters. Although we endeavored to utilize a function similar to GridSearch or experiment with Term Frequency-Inverse Document Frequency (TF-IDF), our implementation in the model was not successful due to our limited comprehension of their mechanisms. However, enhancing this implementation is pivotal to address the model's shortcomings and to understand its potential for what we are trying to achieve. In order to achieve this, it's crucial to take a holistic approach towards understanding how the main goal can be realized through the application of particular tools and strategies. This necessitates a strong emphasis on enhancing skills in Natural Language Processing and programming languages.

Appendix A Topic Labeling for LDA Model

Topic	System	Problems	Solutions
’1+8+4 eğitim sistemi’	1		
’17 mayıs 2019 eğitim sisteminin tekrar değişmesi’	1		
’1+3+3+3+2 eğitim sistemi’	1		
’1+5+4+3 eğitim sistemi’	1		
’5+7 yeni eğitim sistemi’	1		
’9 şubat 2020 yeni eğitim sistemi müjdesi’	1		
’evrim teorisinin anlatılmadığı eğitim sistemi’		1	
’eğitim sisteminde eksikliği en çok hissedilen ders’		1	
’eğitim sistemini düzeltmek için öneriler’			1
’eğitim sisteminin düşünen birey yetiştirmemesi’		1	
’finlandiya eğitim sistemi’	1		
’laik eğitim sistemi gençleri mankurtlaştırdı’		1	
’mehmet demirkolún eğitim sistemi tespiti’		1	1
’sen bir balıksın ve eğitim sistemi senden ağaca’	1		

'türk eğitim sisteminin sorunları'		1	
'tembel olup tüm suçu eğitim sistemine atmak'	1		
'trńin finlandiya eğitim sistemini kopyalamaması'		1	
'türk eğitim sisteminin en büyük sorunu'		1	
'türk eğitim sisteminin iflas etmesi'		1	
'türkiye eğitim sisteminin başarısız olma nedeni'		1	
'türkiye'de eğitim sisteminin başarısız olma nedeni'		1	
'türkiye'deki eğitim sistemi'	1	1	1
'türkiye'deki eğitim sisteminin ana sorunu'		1	
'türkiye'deki eğitim sisteminin en başarısız dersi'		1	
'yeni eğitim sistemi için bir tavsiye bırak'			1
'yurtdışından eğitim sistemimizi soran ülkeler var'	1		
'zorla flüt çaldıran eğitim sistemi'	1		
'eğitim sistemimiz iyi olmasa süper güç olur muyduk'	1		
'sınıf defterinin türk eğitim sistemindeki yeri'	1		

'zekileri mühendisliğe yönlendiren eğitim sistemi'	1	1	
'hüseyin nihal atsız'ın eğitim sistemi önerisi'			1
'finlandiya eğitim sistemi'	1		
'hayalindeki eğitim sistemi'	1		1
'türk milli eğitim sisteminde militarizm'		1	
'türk eğitim sistemini kurtarma planı'	1		1
'internet varken eğitim sistemi diye ağlayan keko'		1	
'çarpım tablosunu bile ezberleten eğitim sistemi'		1	
'eğitim sistemini olumsuz yönde etkileyen faktör'		1	
'eğitim sistemi ile ilgili fikirler'	1		
'eğitim sisteminde ak parti devrimleri'	1		
'ezberci eğitim sistemi'	1		
'türkiye'de eğitim sistemi deyince akla gelenler'	1		
'eğitim sistemi'	1		
'eğitim sistemimiz dinimize göre ayarlanmak zorunda'	1		
'türkiye'de eğitim sistemi son derece iyi'	1		
'milli eğitim sisteminin amacı'	1		

’ideal eğitim sistemi’			1
’kemalist eğitim sisteminin zararları’		1	
’elmastan değerli bir eğitim sistemimiz var’	1		
’karma eğitim sistemi zinaya zemin hazırlamaktadır’	1		
’türk eğitim sisteminin iflası’		1	
’kemalist eğitim sistemi’	1		
’finlandiya eğitim sisteminde devrim’	1		
’japon bakan 4+4+4 eğitim sistemine gıpta etti’	1		
’türkiye’nin eğitim sistemi’	1		
’türk eğitim sisteminde eksik en önemli ders’		1	
’amerikan eğitim sistemi’	1		
’8 dil öğreten norveç eğitim sistemi’	1		
’ldp’nin eğitim sistemi önerisi’			1
’türkiye eğitim sistemi çürüyor ve çöküyor’	1	1	
’örgün eğitim sisteminin uzaktan yapılması’	1		
’derste su içirtmeyen örgün eğitim sistemi’	1		
’türkiye’nin geri kalma sebebi eğitim sistemidir’		1	
’eğitim sisteminin eksiklikleri’		1	

'eğitim sistemi yapboz değil herkes memnun'	1		
'4+1+3+4 eğitim sistemi'	1		

Table 1: Topic Labeling for the Models

Appendix B Original Versions of the Outputs

B.0.1 System

(1, 'yok, iyi, etmek, sadece, sade, kaynak, olmak, devlet, iş, para')

(2, 'düz, din, etmek, el, ev, düzmek, yok, ilk, sadece, sade')

(3, 'sistemi, not, sey, basari, found, temel, laik, ingilizce, sonuç, matematik')

(4, 'la, etmek, saat, sistemidir, bi, örnek, para, sistemi, test, sonuç')

(5, 'hak, ilk, sonuç, etmek, nesil, iyi, yaş, yöntem, türk, büyük')

(6, 'olmak, sadece, sade, yok, etmek, büyük, bok, bi, iyi, para')

(7, 'yok, bi, iyi, özel, olmak, etmek, gerçek, üzerine, öğretim, sistemi')

(8, 'yok, reform, sistemi, iyi, icin, turk, olmak, yoktur, eksik, türk')

(9, 'sade, nesil, sistemdir, sadece, hak, ezber, olmak, tek, genç, üzerine')

(10, 'büyük, sadece, sade, etmek, zaman, sonuç, olmak, yer, örnek, değer')

(11, 'kalite, toplam, yok, iş, birim, ay, milli, mill, millî, mil')

(12, 'etmek, yok, iş, olmak, hak, hoca, ne, dünya, zaman, anlamak')

(13, 'iyi, yok, hedef, olmak, sistemi, etmek, uzak, son, güzel, yer')

(14, 'özel, fen, doğru, hatip, ortalama, temel, yanlış, test, imam, devlet')

(15, 'etmek, olmak, iyi, mezun, büyük, ilk, devlet, edi, orta, gereken')

(16, 'iyi, bölüm, okumak, yok, zaman, iş, insa, sade, aynı, sadece')

(17, 'büyük, orta, tane, etmek, tek, devlet, güneş, resim, olmak, zaman')

(18, 'etmek, olmak, yok, iyi, ilk, yer, meslek, adam, iş, okumak')

(19, 'etmek, olmak, iyi, zaman, son, iş, yok, özel, dünya, konu')

(20, 'tarih, yok, dünya, merak, etmek, gün, başka, var, olmak, varmak')

B.0.2 Problems

(1, 'you, yama, ehliyet, are, ten, ara, yeni, garip, ad, duygu')

(2, 'yüksek, devlet, özel, devam, deva, zaman, el, olacak, okumak, para')

(3, 'üzerine, kurulu, üzer, üzeri, kuru, cevap, ezber, kurul, tek, türkiyede')

(4, 'bi, al, hafta, bilmek, hoca, anadolu, nesil, ge, ilk, anlamak')

(5, 'siyaset, eksik, tarih, hak, tek, ömür, fransız, lan, bi, biri')

(6, 'aynı, gerçek, alan, hayat, insa, anlamak, orta, zaman, okumak, öğrenmek')

(7, 'hoca, bam, resim, sistemidir, resmi, yabanci, pratik, hocanin, hedef, sistemi')

(8, 'cok, icin, degil, sey, ayni, is, ogrenci, olmasi, boyle, oldugu')

(9, 'talep, surekli, hedef, sistemi, yer, arz, birey, takdir, icin, kâfi')

(10, 'balık, olabilir, ge, teknik, gelen, sürekli, ogretmenlerin, ba, ağaç, kuş')

(11, 'orta, yer, toplum, zaman, öğretim, devlet, sorumak, konu, top, sistemi')

(12, 'anne, geçmek, türk, dünya, anlamak, ne, hoca, ödev, baba, icat')

(13, 'meslek, uçak, zeki, disiplin, uzman, yetismis, sonuç, bitene, yöntem, yarayan')

(14, 'din, ahlak, dinmek, ana, olmaması, dil, tarih, dayalı, an, sisteminin')

(15, 'sistemi, olmamasi, bok, sistemdir, şey, suna, sürekli, şeyin, la, başta')

(16, 'desen, mezun, bilim, genç, doğru, in, düzgün, başka, ezber, hukuk')

(17, 'ingilizce, ingiliz, ingilizce, müzik, matematik, türkçe, beden, dil, türk, be')

(18, 'toplum, ne, uzak, amaç, iinsa, devlet, değer, hak, sahip, özgür')

(19, 'hatip, imam, ima, özel, fen, anadolu, fe, orta, sürekli, temel')

(20, 'millî, mil, mill, milli, nesil, kalite, bakan, deva, devam, yönetim')

B.0.3 Solutions

(1, 'özel', 'gerçek', 'olmak', 'bade', 'aynı', 'orta', 'dünya', 'mezun', 'iyi', 'sistemidir')

(2, ’yok’, ’etmek’, ’güneş’, ’bi’, ’olmak’, ’anlamak’, ’meslek’, ’eder’, ’hak’, ’devam’)

(3, ’yok’, ’iyi’, ’etmek’, ’tek’, ’sistemi’, ’büyük’, ’hedef’, ’matematik’, ’el’, ’cok’)

(4, ’etmek’, ’iyi’, ’yeni’, ’iinsa’, ’anne’, ’orta’, ’yok’, ’devlet’, ’ilkokul’, ’sistemi’)

(5, ’etmek’, ’olmak’, ’büyük’, ’iyi’, ’iş’, ’para’, ’konu’, ’yer’, ’üzer’, ’almak’)

(6, ’yok’, ’etmek’, ’özel’, ’anadolu’, ’son’, ’yeni’, ’hoca’, ’olmak’, ’ilk’, ’bok’)

(7, ’saat’, ’türk’, ’sade’, ’sadece’, ’yok’, ’din’, ’zaman’, ’devlet’, ’gerçek’, ’değer’)

(8, ’iyi’, ’olmak’, ’çıkış’, ’sade’, ’sadece’, ’sistemi’, ’sürekli’, ’doğru’, ’kapı’, ’giriş’)

(9, ’dünya’, ’konu’, ’iyi’, ’sade’, ’yüksek’, ’sadece’, ’öğrenmek’, ’olmak’, ’etmek’, ’hakim’)

(10, ’merak’, ’adam’, ’etmek’, ’yazık’, ’başka’, ’yok’, ’yer’, ’icat’, ’at’, ’ada’)

(11, ’nesil’, ’olmak’, ’etmek’, ’zaman’, ’rehber’, ’iyi’, ’gün’, ’tek’, ’eksik’, ’mantık’)

(12, ’kalite’, ’olmak’, ’ilk’, ’kaynak’, ’etmek’, ’iyi’, ’toplam’, ’yok’, ’yer’, ’konu’)

(13, ’iş’, ’iyi’, ’olmak’, ’yok’, ’etmek’, ’yüksek’, ’alan’, ’gerçek’, ’zaman’, ’büyük’)

(14, ’etmek’, ’tahta’, ’olmak’, ’zaman’, ’siyaset’, ’dünya’, ’belli’, ’sistemi’, ’sahip’, ’eğitimin’)

(15, ’reform’, ’hak’, ’devlet’, ’iyi’, ’özel’, ’başka’, ’sade’, ’sadece’, ’okumak’, ’yer’)

(16, ’kalite’, ’olmak’, ’ilk’, ’kaynak’, ’etmek’, ’iyi’, ’toplam’, ’yok’, ’yer’, ’konu’)

(17, ’saat’, ’türk’, ’sade’, ’sadece’, ’yok’, ’din’, ’zaman’, ’devlet’, ’gerçek’, ’değer’)

(18, ’yok’, ’olmak’, ’etmek’, ’yer’, ’nesil’, ’para’, ’sadece’, ’sade’, ’hak’, ’fark’)

(19, ’iş’, ’olmak’, ’etmek’, ’büyük’, ’küçük’, ’yok’, ’değer’, ’iyi’, ’zaman’, ’gelen’)

(20, ’iyi’, ’yok’, ’para’, ’iş’, ’sistemdir’, ’at’, ’aynı’, ’konu’, ’tekrar’, ’olabilir’)

(21, ’iyi’, ’yok’, ’para’, ’iş’, ’sistemdir’, ’at’, ’aynı’, ’konu’, ’tekrar’, ’olabilir’)

(22, ’sistemi’, ’yok’, ’genç’, ’etmek’, ’sadece’, ’sade’, ’olmak’, ’birer’, ’son’, ’dünya’)

(23, ’iş’, ’olmak’, ’etmek’, ’büyük’, ’küçük’, ’yok’, ’değer’, ’iyi’, ’zaman’, ’gelen’)

(24, ’bi’, ’tarih’, ’al’, ’dünya’, ’aynı’, ’nesil’, ’sistemdir’, ’olmak’, ’anne’, ’sadece’)

(25, ’kalite’, ’olmak’, ’ilk’, ’kaynak’, ’etmek’, ’iyi’, ’toplam’, ’yok’, ’yer’, ’konu’)

References

Abu, N. K., and Saral, D. G. (2016). The reasons of academic procrastination tendencies of education faculty students. The Online Journal of New Horizons in Education, 6(1).

Balcı, A. (2008). Türkiye'de Eğitim Yönetiminin Bilimleşme Düzeyi. Kuram Ve Uygulamada Eğitim Yönetimi, 54(54), 181-209.

Can, E. (2015). Qualitative Obstacles in Turkish Education System and Suggestions. Anthropologist, 20, 289-296.

Çelen, F. K., Çelik, A., and Seferoğlu, S. S. (2011). Türk Eğitim Sistemi ve PISA Sonuçları. In Akademik Bilişim'11 - XIII. Akademik Bilişim Konferansı Bildirileri (pp. 765). İnönü Üniversitesi, Malatya.

Çelik, E., Dal, D., and Aydin, T. (2021). Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması. Avrupa Bilim Ve Teknoloji Dergisi(27), 880-889. https://doi.org/10.31590/ejosat.905259

Cross Validated. (2018, January 21). Inferring the number of topics for gensim's LDA - perplexity, CM, AIC, and BIC [Online forum post].

Güngör, T. A., and Pehlivan, O. (2021). Kapsayıcı eğitimin Türk eğitim sistemindeki yeri ve önemi. Studies in Educational Research and Development, 5(1), 48-71.

Kartal, S. E., Kartal, S. E., and Özdemir, R. Y. V. T. Y. (2015). Öğretmenlerde Algılanan Örgütsel Destek Düzeyi ile Yaşam Memnuniyeti Arasındaki İlişkinin

İncelenmesi. Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi(24), 477-504. https://doi.org/10.14582/DUZG

Neyişçi, N., Turabik, T., Gün, F., Kısa, N. (2020). Öğretmen Adaylarının Perspektifinden Türk Eğitim Sisteminin Sorunları ve Olası Çözüm Önerileri. Kastamonu Education Journal, 28(6), 2257-2270. https://doi.org/10.24106/kefdergi.833516

Ömeroğlu, Ç., Daz, Y., Budak, M.B., Üçgül, D., Sağıroğlu, F., and Konuk, Y. (2023). Avrupa Birliği Kurucu Ülkeleri ve Türkiye Okul Öncesi Eğitim Sistemleri. Social Sciences Studies Journal (SSSJournal).

Özer, M. (2020). Türkiye'de COVID-19 Salgını Sürecinde Milli Eğitim Bakanlığı Tarafından Atılan Politika Adımları. Kastamonu Education Journal, 28(3), 1124-1129. https://doi.org/10.24106/kefdergi.722280

Saylık, A., Saylık, N., and Sağlam, A. (2021). Eğitimcilerin Gözünden Türk Eğitim Sistemi: Bir Metafor Çalışması. Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, 18(1), 522-546. https://doi.org/10.33711/yyuefd.919424

Şener, G. (2018). Türkiye'de yaşanan eğitim sorunlarına güncel bir bakış. Millî Eğitim, (218), 187.

Stack Overflow. (2020, July 22). Gensim lda gives negative log-perplexity value - is it normal and how can i interpret it? [Online forum post].

Telci, A. M. (2011). Amerika Birleşik Devletleri eğitim sistemi ile Türk eğitim sisteminde denetim alt sistemlerinin karşılaştırılması (Doktora tezi, Ankara Üniversitesi /

Eğitim Bilimleri Enstitüsü).

Tertemiz, N., and Koç Aytekin, G. (2018). PİSA SONUÇLARININ 2003-2015 EĞİTİM SİSTEMİ VE EKONOMİK GÖSTERGELER KAPSAMINDA İNCELENMESİ: TÜRKİYE VE GÜNEY KORE ÖRNEĞİ. Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi, 19(1), 103-128. https://doi.org/10.29299/kefad.2018.19.004

Terzi, L. (2014). Reframing inclusive education: educational equality as capability equality. Cambridge Journal of Education, 44(4), 479–493. https://doi.org/10.1080/0305764X.2014.960911

Zelner, J., Broen, K., and August, E. (2022). A guide to backward paper writing for the data sciences. Patterns, 3(3), 100423. https://doi.org/10.1016/j.patter.2021.100423