Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi
Pamukkale University Journal of Engineering Sciences

System of automatic scientific article summarization in Turkish

Türkçe otomatik bilimsel makale özetleme Sistemi

Nazan KEMALOĞLU ALAGÖZ[1], Ecir Uğur KÜÇÜKSİLLE[2]*

[1]Department of Computer Technologies, Uluborlu Selahattin Karasoy Vocational School, Applied Sciences University of Isparta, Isparta, Turkey.
nazanalagoz@isparta.edu.tr
[2]Department of Computer Engineering, Faculty of Engineering, Suleyman Demirel Universty, Isparta, Turkey.
ecirkucuksille@sdu.edu.tr

Received/Geliş Tarihi: 31.01.2023 Revision/Düzeltme Tarihi: 05.07.2023 doi: 10.5505/pajes.2023.77905
Accepted/Kabul Tarihi: 14.08.2023 Research Article/Araştırma Makalesi

Abstract

The widespread use of the internet today, along with the rapidly increasing information, has brought along great information pollution. it has become a big problem for internet users to obtain meaningful data from this large and noisy data. Text summarization, which is generally used on texts obtained from digital media, has also been used for summarizing scientific articles in different fields. in this study, a scientific text summary study was carried out to be used on Turkish articles written in the field of informatics. A large Turkish informatics Literature dataset was created with the articles collected from Dergipark. in addition to the text pre-processing studies available in the literature on this dataset, a new original pre-processing function has been developed by the scientific article format. While summarizing, Deep Belief Networks (DBN), which has an increasing use in the field of natural language processing in the literature, has been used. To measure the performance of the developed system, reference summaries were created with the BERT algorithm, which is a pre-trained natural language processing model. After the scientific articles were summarized with BERT and Deep Belief Networks, the abstracts were compared with BERT Score and BART Score, a specialized comparison metric of the BERT Model. The results showed that the developed Turkish informatics Literature Summarization Method constitutes a summary of a scientific article with 0.78 F-Score and 0.68 BART Score in the BERT Score metric.

Keywords: Turkish Natural Language Processing, Automatic Text Summarization, Deep Belief Networks, BERT Score, BART Score.

Öz

Günümüzde internet kullanımının yaygınlaşması, hızla artan bilgi ile birlikte büyük bir bilgi kirliliğini de beraberinde getirmiştir. bu büyük ve gürültülü verilerden anlamlı veriler elde etmek internet kullanıcıları için büyük bir sorun haline gelmiştir. Genellikle dijital ortamlardan elde edilen metinler üzerinde kullanılan metin özetleme, farklı alanlardaki bilimsel makalelerin özetlenmesinde de kullanılmaktadır. Bu çalışmada bilişim alanında yazılmış Türkçe makaleler üzerinde kullanılmak üzere bilimsel metin özet çalışması yapılmıştır. Dergipark'tan toplanan makalelerle geniş bir Türk Bilişim Literatürü veri seti oluşturulmuştur. Bu veri seti üzerinde literatürde mevcut olan metin ön işleme çalışmalarına ek olarak bilimsel makale formatı ile yeni özgün bir ön işleme fonksiyonu geliştirilmiştir. Özetleme yapılırken literatürde doğal dil işleme alanında kullanımı giderek artan Deep Belief Networks (DBN) kullanılmıştır. Geliştirilen sistemin performansını ölçmek için önceden eğitilmiş bir doğal dil işleme modeli olan BERT algoritması ile referans özetleri oluşturulmuştur. Bilimsel makaleler BERT ve Deep Belief Networks ile özetlendikten sonra, özetler BERT Puanı ve BERT Modeli'nin özel bir karşılaştırma metriği olan BART Puanı ile karşılaştırıldı. Elde edilen sonuçlar, geliştirilen Türk Bilişim Literatür Özetleme Yöntemi'nin BERT Puanı metriğinde 0.78 F-Puan ve 0.68 BART Puanı ile bilimsel bir makalenin özetini oluşturduğunu göstermiştir.

Anahtar kelimeler: Türkçe Doğal Dil İşleme, Otomatik Metin Özetleme, Derin İnanç Ağları, BERT Skor, BART Skor.

1 Introduction

The advancement of today's technologies and the widespread use of the internet have brought with it a large data pile. The importance of obtaining a meaningful output from this data stack is increasing. Being able to obtain the most important parts of the data stack, the main information that the data wants to tell, in short, the most summary information that can represent the data will both increase the understandability of that data and save a great deal of time. The abstract is the text extracted from one or more documents and containing the basic information in its source [1] . AutoText Summarization, on the other hand, is to obtain a summary containing the basic information of the text as an output from a text given as input to the computer [2]. in automatic text summarization, the summary extracted from a single document is considered as a single document; The summary obtained from more than one document is called a multi-document summary [2;1]. in multi-document summarization, many documents can be processed, and summaries can be obtained from documents in different formats [3]. According to the type of abstract obtained in text summarization, there are two abstracts, extractive and interpretive. in inferential summarization, a summary is created by selecting the sentences that will best represent the document from among the documents. in inferential summarization, the structure of the sentences is not distorted, the places of the words are not changed. in this summary, sentence selection is made by weighting sentences with statistical methods, intuitive inferences, or hybrid systems where these two are used together [4]. in the summary based on interpretation, the sentences on the document are interpreted and reconstructed. For summary based on comments; A rich source of vocabulary and grammar is required [5].

The document to be summarized in automatic text summarization; it can be texts created by a certain editor, such as an article, news, scientific article, or texts that reflect the views and thoughts of users collected from various social media platforms. Most of the scientific information is found in

*Corresponding author/Yazışılan Yazar

scientific articles, and with the expansion of research areas, it can be quite difficult for academics to find articles that fit their interests. Even query-based searches for some domains return many related articles that exceed human processing capabilities. An automatic summary of these articles will help reduce the time required to review them fully and to get the gist of the information they contain [6].

When the literature is examined, it is seen that the Natural Language Processing studies on Turkish concentrate on news texts. Studies on scientific publication remained in keyword extraction studies. When scientific article summary studies are examined, it is seen that these studies are mainly conducted in English. When examining the literature, it is seen that the studies for Automatic Text Summation in Turkish Natural Language Processing are done on certain texts and the lack of a comprehensive dataset for Turkish text summarization. The contributions of this study to the literature are listed below[1].

- Turkish Scientific Article Dataset was created by us.
- For a scientific article, a specialized pre-processing function has been created in the Turkish language.
- A summary system based on sentence selection with 6 features of a scientific article was developed using Deep Belief Networks.
- For the test of the developed system, the BERT Score metric was used instead of the classical ROUGE metrics and the performance of the created system was demonstrated.

2 Related works

The first study in the field of automatic text summarization was made by H.P.Lunn in 1958 [7]. in this study, the frequency of using words in sentences was used to summarize, and it was suggested that the words with the highest frequency of use were included in the sentences where the most important information about that text was given. in this approach, it was stated that a reader focused on some basic words related to the text, and sentences containing these words were chosen as the main reason.

The first known study for Turkish in the field of summarization belongs to Altan [8]. in this study, 50 documents were used. in developing a system for summarizing, Altan used statistical methods in this system. Documents with these methods; Paragraphs, sentences, words were separated, and summary data were created based on predetermined weight values. in this developed system, options for selecting documents for the user, determining the summary length, and deciding whether to add the summary to the database are also presented.

Guran applied the non-negative matrix decomposition technique on 100 Turkish news datasets. in the study, Guran proposed a new preprocessing step that improves performance. At this stage, the sequential words in the text were determined by using the "Turkish Wikipedia" link structure, and the positive effects of this determination on the method were shown [2].

Working in the field of biostatistics, Yolcular aimed that health care personnel access the information they need by using the abstracts of the articles in the Pubmed literature database. For this purpose, a web interface has been developed by using an inferential summarization method, by selecting medical terms and sentences containing statistical data [9].

Kim and colleagues have prepared a computer science paper dataset from ArXiv. The paragraphs of the entire text in the article to be summarized were examined. Using the Jaccard analogy, they chose the most informative sentence as the summary of each paragraph. A specialized Auto-encoder RNN system was used for sentence selection [10].

In 2017, Collins et al. performed summarization into a dataset of 10,000 computer science articles. After creating the dataset, they scored each sentence with the traditional sentence weighting method and made an inference-based summary [11].

Wang et al. worked on an interpretative summarization technique with a deep learning approach. Texts written in Chinese are summarized in the system using Reinforcement. While the studied dataset consisted of short texts, it was argued that the developed model could produce summaries with information, consistency, and diversity [12].

Nikolov et al. compared interpretative summarization based on an artificial neural network model with inferential summarization on scientific publications. Although the developed system is quite successful in producing summary titles, it has been seen that it needs improvement in summarizing. in addition, they were emphasized that the abstract sentences extracted were of semantically high quality [13].

Nallapati et al. studied multi-document summarization. A new dataset was created in the study, in which the interpretive summarization system was used. Auto-encoder (AE) artificial neural network algorithm, which is a deep learning method, was used on the dataset. AE is a neural network trained to copy the input it receives to its output. it is seen in the literature that the proposed model improves performance [14].

Sirohi et al. for summarizing abstract text in 2021; They examined various machine learning methods such as Fuzzy Method, Support Vector Machines, Bayesian Summarization. in addition, they developed an Encoder-Decoder-based LSTM architecture for summarization as a new solution. in the study, the encoder-decoder LSTM architecture is used as an attention mechanism. The principle on which the study is based on the priority use of words that have a high relationship with that word in summarizing a word. For example, for the word Hamburger; words like spicy, delicious; it is more important than words like this and that should be given priority. in the research, the attention mechanism was used to solve this process [15].

Lloret, et al. performed a summary study of biomedical publications. The scientific publications used were summarized with both interpretive and inferential summarization methods and the results were compared. The abstracts were evaluated qualitatively and quantitatively by experts. The results show that both summarization methods can extract important information from the scientific article. in addition, it was stated by experts that interpretive summary is more meaningful and understandable [16].

3 Material and method

In this research, a summary study was carried out with Deep Belief Networks, which is a deep learning method, over scientific articles written in the field of informatics. For this purpose, to create a new dataset, 8 journals with Emerging Sources Citation index (ESCI), Science Citation index Expanded (SCI-EXPANDED), and TR index were determined and articles in the field of informatics were collected from these journals.

This dataset was summarized with the DiA model developed for the thesis and reference summaries obtained from the BERT inferential summary were used to evaluate the system performance.

3.1 Dataset

The dataset used for the analysis was composed of articles prepared in the field of informatics obtained from various journals published in Turkey. For this purpose, it was paid attention that the articles to be selected were published in journals with Emerging Sources Citation index (ESCI), Science Citation index Expanded (SCI-EXPANDED), and TR index. Eight journals with Emerging Sources Citation index (ESCI), Science Citation index Expanded (SCI-EXPANDED), and TR index published in Turkish were determined in this way and these journals were filtered according to the informatics articles published on Dergipark. Here, to determine a common year limit, all journals have been scanned since 2017, based on the year of the oldest informatics article available on Dergipark. As a result, a total of 485 articles published in the field of informatics were obtained. Table 1 shows the distribution of the total dataset according to the journals. The dataset created for this study has been published on GitHub and it is available at https://github.com/nazankemaloglu/AutomaticTextSummarization.

Table 1. Informatics Articles Dataset.

Journal	TA	LAP	SAP
Journal of Gazi University Faculty of Engineering and Architecture	77	26	8
Journal of Polytechnic	33	20	5
Journal of Information Technologies	141	25	6
Pamukkale University Journal of Engineering Sciences	41	22	5
Journal of Suleyman Demirel University Institute of Science and Technology	23	15	4
European Journal of Science and Technology	79	24	4
Dokuz Eylul University Faculty of Engineering Journal of Science and Engineering	41	19	5
Fırat University Journal of Engineering Sciences	50	16	5

According to Table 1; TA is Total Article Number, LAP is Longest Article Pages Number and SAP is Shortest Article Pages Number.

3.2 Deep Belief Network

The Restricted Boltzmann Machine (RBM) is a highly successful unsupervised learning in the field of feature extraction from unlabeled data. Thus, an output of the hidden layer in one RBM can be used as the input of the visible layer of another RBM. This process can be considered as more feature extraction than features extracted from the available data. Hinton introduced the Deep Belief Network (DBN) based on RBM in 2006 [17;18]. Deep Belief Networks is a semi-supervised learning model formed by the combination of more than one RBM model [19]. The simple architecture of the Deep Belief Network is given in Figure 1.

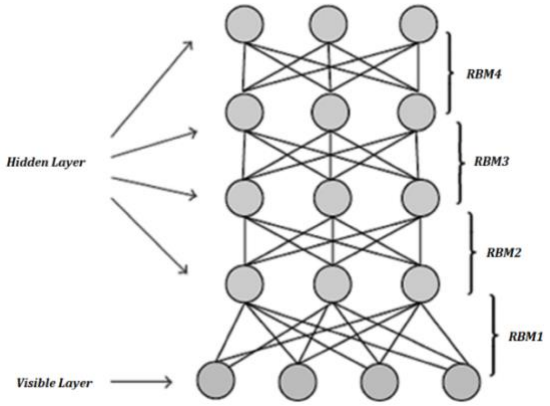

Figure 1. Networks of Deep Beliefs.

As given in Figure 1, DBN training steps are as follows:

1. RBM1 is trained to reconfigure the input.
2. The hidden layer of RBM1 is considered the visible layer of RBM2.
3. With the output of RBM1, RBM2 is trained.
4. This process is repeated until every layer in the network is trained.

3.3 BERT (Bidirectional Encoder Representations from Transformers)

BERT: Bidirectional Encoder Representations from Transformers is a pre-trained language detection model developed by the Google Artificial intelligence division [20;24]. BERT model that produces solutions to Natural Language Processing problems; it has been used by Google so that the search engine can provide healthier results and a high rate of success has been achieved [20]. it is actively used in the current Google search engine system to find the most relevant results between the searched phrase and the results obtained

In the BERT model, it is understood that the word "stand" is related to the physical context of a job, and more useful results are produced [21]. The input representation of the BERT model is given in Figure 2.

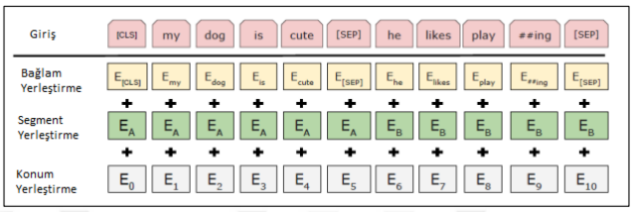

Figure 2. BERT Example Sentence Display [21].

As given in Figure 2 separate markers were added to the beginning and end of each sentence. These are (CLS) which represents the beginning of the sentence and (SEP) which represents the end. This placement is achieved by BERT token generation. Segment placement in the model is used to distinguish the input sentences, while position placement shows the position of the markers in the sequence [22].

In addition to the fact that BERT has models specialized in more than one language since it is open-source, developers can train their language models with different language modules.

To evaluate the performance of the summary method developed for the study, the articles were trained for the BERT

Extractive Summarizer (BERT) method and summaries were obtained [23]. These summaries were used in the performance comparison of the developed summarization model.

3.4 BERT Score

BERT is an open-source, pre-trained natural language processing model developed by Google researchers [24]. BEST SCORE establishes a contextual relationship between two sentences based on cosine similarity [25]. The working principle of the BERT Score is given in Figure 3.

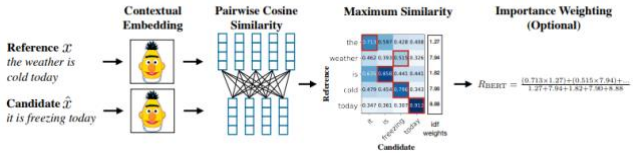

Figure 3. BERT Score Example Chart.

As given in Figure 3, BERT Score uses inter-word contextual insertions between a reference sentence x and a candidate sentence x̂. While calculating here, optionally inverse document frequency (iDF) scores are calculated using predominantly cosine similarity. Here, the BERT algorithm is used for the separation of the main sentence into words [26;25].

The cosine similarity between the two sentences is calculated according to Equation 1.

$$Similarity = \frac{x_i^T \hat{x}_j}{\|x_i\| \|\hat{x}_j\|} \quad (1)$$

When calculating the BERT Score value; For recall, each word in each reference sentence x is matched with each word in the candidate sentence x̂. In addition, for the precision value, each word in each candidate sentence x̂ is matched with each word in the reference sentence x [25].

From this point of view, BERT Score values for reference sentence x and a candidate sentence x̂ are based on Precision (Recall, R) given in Equation 2, Precision (P) given in Equation 3, and precision given in Equation 4, and it is determined by the calculated F (F1 Score) values [25].

$$R_{BERT} = \frac{1}{|x|} \sum_{x_i \in x} \max x_i^T \hat{x}_j \ , (\hat{x}_j \in \hat{x}) \quad (2)$$

$$P_{BERT} = \frac{1}{|\hat{x}|} \sum_{\hat{x}_j \in \hat{x}} \max x_i^T \hat{x}_j \ , (x_i \in x) \quad (3)$$

$$F_{BERT} = 2 \frac{P_{BERT} \cdot R_{BERT}}{P_{BERT} + R_{BERT}} \quad (4)$$

3.5 BART Score

BART is an autoencoder for training sequence-to-seq models. BART uses a standard transformers-based neural machine translation architecture. A standard seq2seq architecture is used with a bidirectional encoder and left-to-right decoder [27].

BART Score is a metric based on the probability that a text is produced from or derived from a given text. BART Score metric is calculated according to Equation 5.

$$BART\ Score = \sum_{t=1}^{m} w_t \ log \ log \ p \ (y_t | y_{<t}, x, \theta) \quad (5)$$

In the above equation, x represents the target text and y represents the reference text, while w is calculated with the Inverse Document Frequency (IDF) [28].

BART Score is developed in Python language and published on GitHub with open-source code. In our study, the BART Score metric was calculated using the code available on GitHub between the BERT Inferential Summarizer and the summary obtained with the DBN.

The developed system consists of 3 parts. In the first stage, the pdf files of the articles are read, and these files are converted into text files by going through the pre-processing stages summarized in Figure 6. In the second stage, the abstract texts were obtained by extracting the attributes from these text files. In the last part, the dataset is summarized with the BERT Inferential Summarization method, which is frequently used in the literature, and the obtained summaries are compared with BERT Score metrics to measure the performance of the developed system. The summary scheme of the developed system is given in Figure 4.

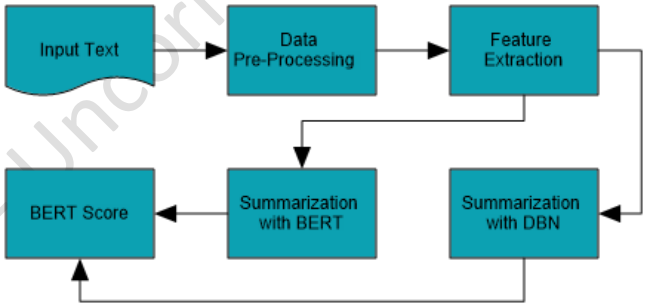

Figure 4. Developed System Diagram.

4 Results

One of the most important steps for natural language processing is preprocessing the text to make it suitable for analysis. The sections of the article examined in this study (Abstract, Bibliography, Material Method, etc.) were examined in the pre-processes according to whether they were effective in the abstract or not. As a result of this review, first, the articles were divided into sections. The main parts that will not be included in the summary have been determined and removed from these sections. Extracted sections; The bibliography consists of English Abstract, Abstract, Author, and information about the journal. A different pre-processing was applied in the remaining text sections. These transactions are deleting the explanations of tables, pictures and equations is the removal of sentences that are considered as references in the text and taken from different articles. The pre-treatments performed are summarized in Figure 5.

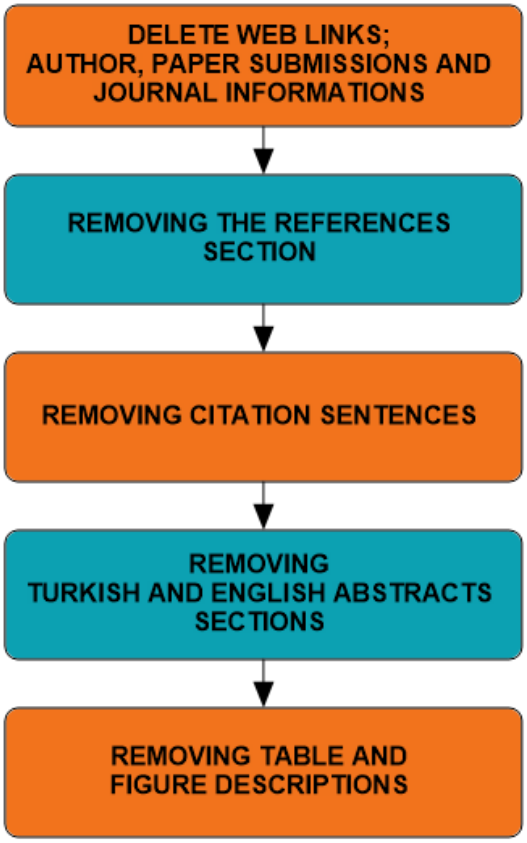

Figure 5. Pre-processes Applied to the Text.

For the operations given in Figure 5 on the articles to be processed, a pre-processing function that can be used on the article was written by using RegEx, NLTK, and LangDetect libraries. This function performs an article-specific preprocessing, considering the specialized text structure of the articles.

Before proceeding to the summary of the pre-processed articles, it is necessary to extract the attributes from the text first. DBN can learn from unlabeled data. In addition to these attributes, sentences with more importance in a scientific article were determined and used as attributes. At this stage, feature extraction with sentence weighting methods was used based on Güran's study in 2013. These attributes:

- Sentence Length

Based on the data that long sentences contain more information than short sentences, sentences are assigned a score according to the number of words.

- Inclusion of Keywords in Sentence

The sentences of scientific articles containing keywords are assigned a score equal to the number of keywords they contain. This feature has been added by us based on the Turkish and English keywords found in a scientific article.

- Passing Status of Numeric Data

It is known that the sentences containing numerical data in scientific articles have high importance in the article. For this reason, sentences containing numeric characters are assigned a score equal to the numeric character they contain.

- Similarity to Title

A score is assigned to the sentence according to the similarity between the article title and the article sentences. This similarity is calculated according to Equation 6 with Cosine Similarity.

$$Cosine(C_i, C_{title}) = \frac{multiply(C_i, C_{title})}{(\|C_i\| \, \|C_{title}\|)} \quad (6)$$

Here C_i represents the i. sentence in the article; C_{title} denotes the title sentence [2].

- Scoring by Word Frequencies

After the frequency of each word in the article was calculated according to Equations 7, 8, and 9, a ranking was made from high to low. Considering 10% of this list, sentences are assigned a score equal to the sum of the frequency values of these words, depending on whether they contain these words.

$$Term\ Frequency\ (TF) = \frac{X}{y} \quad (7)$$

$$Reverse\ Document\ Frequency(IDF) = log\ (\frac{T}{X}) \quad (8)$$

$$TF - IDF = TF * IDF \quad (9)$$

Here; x is number of sentences in which the word occurs in the article, y is total number of sentences in the article, T is Total number of sentences.

- Inclusion of Summarizing Words

In summary, in conclusion, however, besides, in conclusion, etc. Considering that the sentences containing aggregative words such as " are weightier for the articles, a score equal to the number of suffixing words are assigned to the sentences containing these words.

In this study, Deep Belief Networks were used as the summarization method. For the training of the model, features extracted by sentence scoring methods were used. The sentence-feature matrix is constructed in such a way that each sentence has 6 feature vector values. After that, recomputation is performed on this matrix to develop and abstract the feature vectors to form complex features from simple features. This step was used to improve the quality of the abstract. For this, KBM is used, and the sentence-feature matrix is given as an input to KBM. In total, 750 visible and 500 hidden layer CBMs were used. 6 sensors with a learning rate of 0.1 were used in each layer, the data were given to the system in stacks of 4 and a total of 10 epochs were run. RBM is retrained for each new document that needs to be summarized.

The final feature vector values obtained as a result of training the KBMs are summed to form a score against each sentence. The sentences are then sorted by decreasing point values. The most appropriate sentence, the first sentence in this sorted list is chosen as the first sentence of the summary. Then the next sentence selected is the sentence with the highest Jaccard similarity to the first sentence, starting from the top of the sorted list. This process continues iteratively and gradually to select more sentences until a summary limit is reached that will be created with 20% of the sentence number of the introductory text. The sentences are then rearranged according to the order of appearance in the original text, and the final version of the summary text is revealed.

In the summary study, the raw texts were summarized with the BERT Inferential Summarization method to compare the performance of the developed system. A comparison was made between these two abstracts according to BERT Score metrics. When Table 2 and Figure 6 are examined, an average F-Score value of 0.78 was obtained among the summaries obtained from the developed model and the reference summaries obtained from BERT Inferential Summarization, which is accepted in the literature. It is seen that the BART Score metric, which is another metric examined according to Table 2 and Figure 7 and Figure 8, provides a lower overlap than the BERT Score metric with a value of 0.68.

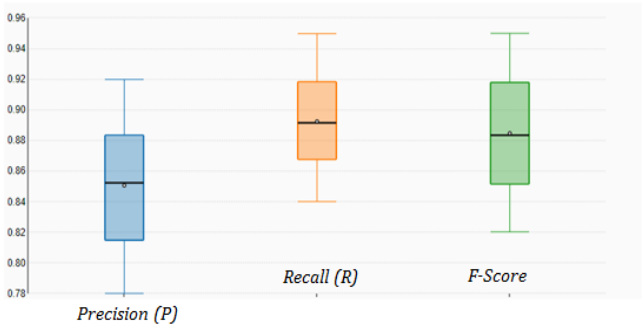

Figure 6. BERT Score Values for the Whole Dataset.

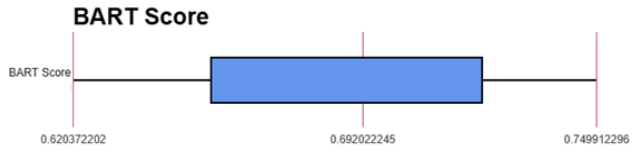

Figure 7. BART Score Values for the Whole Dataset.

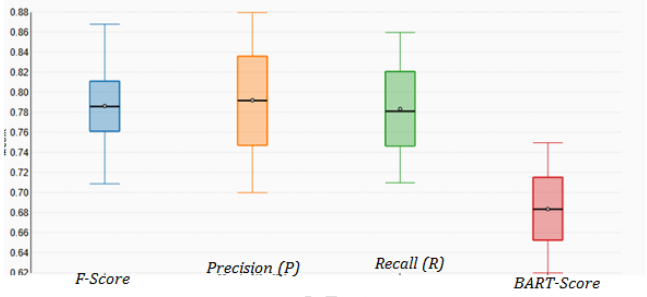

Figure 8. Evaluation Results Values for the Whole Dataset.

5 Conclusion and Future Works

In this study, an original Turkish Scientific Article Dataset was created. Although there are studies that involve summarization in Turkish language, these studies mainly focus on social media [34], news texts [2, 5], user comments [30], and podcasts [31]. There is a tendency in the literature to summarize scientific articles in English. Furthermore, there are studies conducted in languages such as Indonesian [32], Hindi [4], and Chinese [12]. There is a need for a dataset for summarizing scientific publications in Turkish and obtaining these summaries. To fulfill this purpose, a Turkish Scientific Article Dataset was created by obtaining all computer science articles from 8 journals published in Turkish in 2017 that are indexed in the Emerging Sources Citation Index (ESCI), Science Citation Index Expanded (SCI-EXPANDED), and TR Index.

In the preprocessing stage of the dataset, in addition to classical text processing techniques, a new text preprocessing function was created based on the scientific article format by adding new features. For comparing the generated summaries, the BERT Score metric was used based on the ROUGE metrics commonly used in the literature. The reference summaries for comparison were obtained from the BERT Extractive Summarizer, which was developed by Google developers based on the BERT algorithm. When examining the results, the developed system for summarizing Turkish scientific articles achieved an F-Score of 0.78. It was observed that the BART Score metric performed lower than the BERT Score in the evaluation, giving a value of 0.68 for the developed system and the summaries obtained from the BERT Extractive Summarizer.

In the development phase of the Automatic Text Summarization system presented in the study, the lack of specific standardized evaluation criteria for the summary text is the most significant problem. Previous studies show that reference summaries are used for comparing the generated summaries. In addition, evaluation criteria based entirely on intuitive methods are available in the literature. These methods were not found useful due to the diverse and lengthy formats of the scientific articles in the used dataset. In future studies, it is planned to establish a standard for evaluating the generated summary text in automatic summarization systems through a study that can be based on a Turkish dataset. The fundamental principle is to conduct research on the extent to which the extracted summaries represent the source text.

6 Author contribution statements

Author 1: Creation of the idea, design, supply of resources and materials, data collection, analysis, literature review, writing.

Author 2: Creation of the idea, making the design, critical review.

7 Ethics committee approval and conflict of interest statement

"There is no need to obtain ethics committee approval in the prepared article"

"There is no conflict of interest with any person/institution in the prepared article".

Table 2. BERT Score and BART Score Values for the Whole Dataset.

Metrics	Total Number of Dataset(N)	Min	Median	Max	Avg	Std
BERT Precision (P)	485	0.700	0.747	0.792	0.836	0.879
BERT Recall (R)	485	0.710	0.746	0.781	0.821	0.859
BERT F Score (F1)	485	0.708	0.761	0.786	0.811	0.868
BART Score	485	0.620	0.652	0.683	0.715	0.749

8 References

[1] Erhandi B. Text Summarization With Deep Learning. MSc Thesis, Sakarya University, Sakarya, Turkey, 2020.

[2] Guran, A. Automatic Text Summarization System. PhD Thesis, Yıldız Technical University, Istanbul, Turkey, 2013.

[3] Mutlu B. Text Summarization by Hybrid Intelligent System. PhD Thesis, Gazi Universty, Ankara, Turkey, 2020.

[4] Gupta V, Lehal GS. "A Survey of Text Summarization Extractive Techniques". *Journal of Emerging Technologies in Web Intelligence*, 2(3), 258-268, 2010.

[5] Dudak E. Extractive Text Summarization by Gray Wolf Optimization Algorithm and Classification of Abstracts with Deep Learning. MSc Thesis, Duzce Universty, Duzce, Turkey, 2020.

[6] Altmami NI, Menai MEB. "Automatic Summarization of Scientific Articles: A survey". *Journal of King Saud University-Computer and information Sciences*, 34(4), 1011-1028, 2020.

[7] Luhn HP. "The Automatic Creation of Literature Abstracts". *IBM Journal of Research and Development*, 2(2), 159-165, 1958.

[8] Altan Z. "A Turkish automatic text summarization system". *IASTED International Conference Artificiall Intelligence and Applications,* Innsbruck, Austria, 16-18 February 2004.

[9] Yolcular Oguz B. Literature Mining; A Real-Time WEB-based Text Mining Application. PhD Thesis, Akdeniz Universty, Antalya, Turkey, 2016.

[10] Kim M, Singh MD, Lee M. (2016). "Towards abstraction from extraction: Multiple timescale gated recurrent unit for summarization". *1st Workshop on Representation Learning for NLP, Berlin Germany, 11 August 2016.*

[11] Collins E, Augenstein I, Riedel S. "A supervised approach to extractive summarisation of scientific papers". *21st Conference on Computational Natural Language Leraning, Vancouver, Canada,* 2017.

[12] Wang L, Yao J, Tao Y, Zhong L, Liu W, Du Q. "A reinforced topic-aware convolutional sequence-to- sequence model for abstractive text summarization". *Internatioanl Joint Conference on Artificial Intelligence*, Stockholm, Sweeden, 13-19 July 2018."

[13] Nikolov, N. i., Pfeiffer, M., & Hahnloser, R. H. (2018). "Data-driven Summarization of Scientific Articles". *Language Resources and Evaluation Conference,* Miyazaki, Japan, 7-12 May 2018.

[14] Nallapati R, Zhou B, Gulcehre C, Xiang B. "Abstractive text summarization using sequence-to-sequence rnns and beyond". *The SIGNLL Conference on Computational Natural Language Learning*, Berlin, Germany, 11-12 August 2016.

[15] Sirohi NK, Bansal M, Rajan SN. "Recent Approaches for Text Summarization Using Machine Learning & LSTM0". *Journal on Big Data*, 3(1), 35-47, 2021.

[16] Lloret E, Roma-Ferri MT, Palomar M. "Compendium: A Text Summarization System for Generating Abstracts of Research Papers". *Data Knowledge Engineering*, 88, 164–175, 2013.

[17] Hinton GE, Osindero S, Teh YW. (2006). "A Fast Learning Algorithm for Deep Belief Nets". *Neural Computation*, 18(7), 1527-1554, 2006.

[18] Hua Y, Guo J, Zhao H. "Deep belief networks and deep learning". I*nternational Conference on Intelligent Computing and internet of Things*, Harbin, China, 17-18 January 2015.

[19] Savas BK, Becerikli YA. "Deep Learning Approach to Driver Fatigue Detection via Mouth State Analyses and Yawning Detection. *Journal of Computer Engineering,* 23(3), 24-30, 2021.

[20] Sar KT. Time Based Sentiment Analysis Using Artificial Neural Networks and Bert Language Model: Exploring Comments on Whatsapp's New Privacy Policy, MSc Thesis, Dokuz Eylul Universty, Izmir, Turkey,2021.

[21] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[22] Ozkan B. "Dialog Intent Classification Using NLP Methods". MSc Thesis, Bahcesehir Universty, Istanbul, Turkey, 2021.

[23] Arxiv. "Leveraging BERT for Extractive Text Summarization on Lectures". arXiv preprint arXiv:1906.04165. (20.11.2021).

[24] Arxiv. "Bert: Pre-training of deep bidirectional transformers for language understanding". arXiv preprint arXiv:1810.04805. (20.11.2021).

[25] Arxiv. "Bertscore: Evaluating text generation with bert". arXiv preprint arXiv:1904.09675. (20.11.2021).

[26] Arxiv. "Google's neural machine translation system: Bridging the gap between human and machine translation". arXiv preprint arXiv:1609.08144. (20.11.2021).

[27] Arxiv. "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension". arXiv preprint arXiv:1910.13461. (23.10.2021).

[28] Yuan W, Neubig G, Liu P. "Bartscore: Evaluating generated text as text generation". *Conference on Neural Information Processing Systems,* Virtual Conference, 6-14 December 2021.

[29] Github. "An Approach to Automatic Trending Tweet Summarization". https://github.com/yuva29/twitter-trends-summarizer (04/20/2018)

[30] Das SJ, Murakami R, Chakraborty B. "Development of a Two-step LDA Based Aspect Extraction Technique for Review Summarization". International Journal of Applied Science and Engineering, 18(1), 1-18, 2021.

[31] Karlbom H, Clifton A. "Abstractive Podcast Summarization using BART with Longformer Attention". *Text Retrieval Conference*, 2021.

[32] Slamet, C., Atmadja, A. R., Maylawati, D. S., Lestari, R. S., Darmalaksana, W., & Ramdhani, M. A. (2018). "Automated text summarization for indonesian article using vector space model. The 2nd Annual Applied Science and Engineering Conference (AASEC 2017), Bandung, Indonesia, 24 August 2017.

Appendix A

Four Article Summary of System

SUMMARY 1: "Akan Veri Kümeleme Teknikleri Üzerine Bir Derleme", (DOI: 10.31590/ejosat.446019)

Teknolojinin hızlı bir şekilde gelişmesine paralel olarak bilgisayar, akıllı telefon veya sensor gibi veri üreten cihazların kullanımı yaygınlaşmıştır.Bu tür cihazlar kullanılarak "Büyük Veri" olarak tanımlanan inanılmaz bir veri miktarı üretilmektedir.Bunların başında verileri bir yere kaydetme gelmektedir.Çünkü çok büyük veri miktarını kayıt altına almak için gereken kaynak ihtiyacı da çok büyüktür.Örneğin banka gibi finansal kuruluşlar için analizlerin anlık yapılabilmesine ihtiyaç vardır.Kaç sınıf olacağını ya kullanıcıdan girdi olarak alır ya da rastgele belirleyerek işlem yapmakta ve sonuçta bu girdilere göre her zaman doğruluğu kesin olmayan bir kümeleme yapmaktadır.Bu çalışmada akan veri kümeleme alanında yapılan çalışmalar derlenerek bu alanda çalışmayı düşünen araştırmacılara ışık tutmak hedeflenmektedir. Adı geçen uygulamalarda yüksek hızda anlık veri akmaktadır.Başı ve sonu önceden bilinemez.Dolayısıyla verinin geneli hakkında bilgi sahibi olma imkânı yoktur.Geliştirilecek yöntemin kaynakları etkin kullanması ve tatmin edici sonuçları makul bir zamanda sunması ihtiyacı bunlardan bazılarıdır.Bu nedenle geliştirilecek yöntemlerin bu problemleri göz önünde bulundurması gerekir. Dinamizm: Veri dinamiktir, her an değişir.Kullanıcının istediği her an küme sonuçlarını sunabilmesi gerekir.Nesnelerin interneti konusunun yaygınlaştığı günümüzde uygulama alanlarının daha da artacağını söylemek mümkündür.Yukarıda da bahsettiğimiz gibi teknolojinin gelişmesine paralel olarak internet ortamına aktarılmış olan veri miktarı üssel olarak artmaktadır.Bu devasa verilerin anlamlı bilgiye dönüştürülmesine olan talep ve yeni ihtiyaçlar bu alanda yapılacak çalışmalara ilham vermektedir.Çünkü çalışmaların odaklandığı sapan veri tespiti veya çok boyutluluk gibi problemleri aynı anda sağlayan gerçek verileri bulmak çok zordur.Bu nedenle bu alanda çalışma yapan araştırmacılar genelde kendi verilerini üretmektedir.Başlıca veri özetleme yöntemleri şunlardır: Rastgele Örnekleme (Random Sampling): En basit özetleme yaklaşımıdır.Bu özetleme yaklaşımında belli aralıklarla örnek alınır.Hafıza açısından oldukça tasarrufludur, sadece w kadar hafızaya gereksinim duyar.Bu yaklaşımda veriler kova (bucket) denilen parçalara bölünür.Kovaların boyutu gerekli hafıza miktarını belirler.Dezavantajı geriye dönük işlem yapamamasıdır.Histogram örnekleme örneği Çok Çözünürlüklü Metodlar (Multiresolution Methods): Büyük veri miktarı ile baş etmek için geliştirilmiş bir yöntemdir.Bu yaklaşımda veri miktarını azaltmak için parçala ve fethet yaklaşımı izlenir.Bu yaklaşımın en önemli avantajı veriyi yönetme anlamında doğruluk ve kaynak açısından optimum sonuç vermesidir.Ayrıca veriyi farklı açılardan anlamaya yardımcı olur.Örnek olarak dengeli ikili ağaçlarını ele alırsak kökten yapraklara doğru indikçe her adımda sonuç detaylandırılmış olur.Bu yaklaşım da buna benzemektedir.İki çeşit çok çözünürlüklü metot vardır.Taslak (Sketches): En doğruya yakın bir cevabı elde etmeyi garanti eden bir yaklaşımdır.Veriye ait sıklık momenti (frequency moment) sketch denilen özetler ile elde edilir.Rastgele örnekleme ve taslak yaklaşımlarının birleşiminden oluşur.Purity Testi, F-Score, Accuracy, Rand index(RI), Adjusted Rand index(ARI) ve Silhouette index bu alanda kullanılan başlıca yöntemlerdir.Akan veri kümeleme modelini değerlendirirken bu parametrelerden sadece birini kullanmak başarıyı tam olarak ölçmek adına yeterli değildir.Bu nedenle bu parametrelerden birkaç tanesi kullanılmaktadır.Örneğin Purity-ARI, Purity-Accuracy-F-Score veya Purity-ARI-Silhouette İndex parametrelerinin kullanımı oldukça yaygındır.Temel anlamda kümeleme başarısı değerlendirme yöntemlerini iki gruba ayırabiliriz: Dahili ve harici yöntemler.Söz konusu bu benzerlik kriteri çoğu zaman uzaklıktır.Silhouette index ve SSE (Sum of Squarred Errors - Hataların Kareleri Toplamı), başlıca dahili küme başarısı değerlendirme yöntemleridir.Bu uzaklıklardan ilki verinin bulunduğu kümeye ait diğer verilere olan uzaklıkların ortalamasıdır.Verilerin hata paylarının karelerinin toplamını bulur.Yani veri doğruya ne kadar yakın olursa hata paya o kadar azalır.Purity önerilen modelin yaptığı kümeleme yaklaşımının saflık derecesini hesaplar.Her küme için içerisinde barındırdığı verilerden sayısı en fazla olan verilerin toplam veri sayısına oranıdır.Bu işlemi tüm kümeler için yapar ve elde edilen sonuçların tamamını toplar.Bu toplamın toplam veri sayısına oranı Purity'dir.Sonrasında bu yapılan işlemi ikili olarak karşılaştırır.Aynı mı yoksa farklı mı olduğunu bir tabloda tutar.Bu tablo üzerinden RI değerini hesaplar.Oysa ki harici yöntemlerde verilen etiketin gerçek küme etiketi ile uyuşmaması, sonucu doğrudan etkiler.Zaman karmaşılığı açısından bakıldığı zaman ARI, Jaccard index, Fowlkes & Mallows ve Silhouette index gibi kümeleme başarısı değerlendirme yöntemlerinin zaman karmaşıklığı diğer yöntemlerden daha yüksektir.Bu tür yaklaşımlarda küme şekli yuvarlaktır.Bu yaklaşımlarda genelde sonuç gürültü ve sapan veriden etkilenir.Küme birleştirme (merge) ve küme bölme (split) işlemi sürekli yapılır.Hiyerarşik kümelemenin başarısını arttırmak için diğer yöntemlerle birleştirilmiş yaklaşımlar da mevcuttur.Grid tabanlı kümeleme, verinin dağılımından bağımsızdır.Grid tabanlı yaklaşımların yoğunluk tabanlı yaklaşımlarla birleştirildiği çok sayıda çalışma da mevcuttur.Yoğunluk tabanlı kümeleme, yoğunluğu baz alarak kümeleme yapar.Bu yaklaşımlarda kümeler verinin yoğunlaştığı alanlarda yoğunlaşır.Temel yaklaşımı kümeyi verinin yoğunlaştığı yerlere doğru genişletmeye dayanır.Böl ve fethet yaklaşımına dayanır.İlk aşamada data stream kova (bucket) denilen parçalara bölünür ve her kova için k-median uygulanarak k tane küme bulunur.Veriye ait özetleme istatistikleri belleğe kaydedilir.Bu durum, kullanıcıya

zaman açısından esneklik sağlar.Bu algoritma dairesel olmayan şekilleri bulmakta çok verimli değildir.Bu yaklaşım temel anlamda verimli bellek kullanımı ile ön plana çıkmaktadır.StreamSamp'te sabit boyutlu bellek kullanımı benimsenmiştir.İki çeşit hafızalama yoluna gidilmektedir.Kısa boyutlu ve daha uzun boyutlu hafızalama.Online kısım mikro-kümeleri bulmak için kullanılır.mikro-kümeler hiyerarşik yapıda birleştirilir.Sistem kendinden uyarlamalıdır.CluStream algoritmasının gelişmiş bir versiyonudur.ClueStream algoritmasına göre avantajı çok boyutluluğu desteklemesidir.Çok boyutlu verilerde boyutların bir alt kümesini alır.Nitelik sayısı güncellenebilmektedir.Yeni gelen verilere daha fazla önem verilmesi gereken uygulamalar için uygundur.Çok boyutluluğu desteklemesine rağmen çok boyutluluk konusunda optimum bir sonuç verememektedir.Küme şekillerini belirlemekte problem çıkmaktadır.İyi performans için verinin tamamı hakkında yeterli bilgiye sahip olmak gerekir.Gridler noktalar halinde haritalanır ve depth first search algoritması uygulanarak kümeler oluşturulur.Kümeleri bir grafın parçaları gibi birbirine bağlar.rDenStream algoritmasının en önemli farkı atılan verilere ikinci bir şans vermesidir.Fading cluster yapısını ve histogram kullanır.Eski gridlerin ağırlığını azaltmak için zamana bağlı olarak ağırlık azaltma (fading) fonksiyonu kullanır.Eğer belirlenen bir gridin yoğunluk değeri eşik değerinin altına düşerse ve yeni veriler eklenmez ise silinir.Nitelik sayısı (dimension) arttıkça grid sayısı üssel artar, bu nedenle çok boyutlu problemler için uygun değildir.Amaç çok boyutlu veriler için performansı arttırmak.Boyut indirgeme işlemi yapılır.Bunu yaparken de sonuçla en fazla alakası olan nitelikler seçilir.SE-Stream algoritmasında bunlar optimize edilir.n-boyutlu veriyi gridlere ayırır ve özvektör (eigenvector) ile update eder.Özvektör grid merkezinin koordinatlarını, gridin en son güncellendiği zamanı, gridin grid merkezinden silindiği zamanı ve en son grid yoğunluğu gibi veri gruplarını içerir.Özvektör ile gridin yoğun mu yoksa seyrek bir grid mi olduğuna karar verir.Kategorik niteliklerde belirsizliği ortadan kaldırmak için iki objenin uzaklık fonksiyonunu olasılık dağılımı ile kullanır.K üme yapısında değişiklik yapılıp yapılmayacağına adı geçen uzaklık fonksiyonu ile karar verir.HDDStream üç aşamadan oluşur.HDDStream algoritmasının HPStream algoritmasından farkı küme sayısının zamana bağlı olarak değişmesidir.CL-Ant ve CL-AntInc algoritmalarında veriler karınca olarak temsil edilmektedir.HSDStream üç aşamadan oluşmaktadır.Yapılan deneysel çalışmalar HSDStream algoritmasının HDDStream algoritmasından daha iyi sonuçlar verdiğini ortaya koymuştur.CODAS'ın getirdiği en önemli yenilik tamamen online çalışmasıdır.Bunu gerçekleştirmek için kümelemede hyper-spherical mikro-küme denilen yapıyı kullanır.k-means ve k-medoid gibi yaklaşımların aksine herhangi bir varsayımda bulunmaz.Skeleton denilen parçalar verinin yoğunluğa göre ağırlıklandırılmış şeklidir.Mevcut yaklaşımların çoğundan daha iyi sonuç verdiği tespit edilmiştir.Bu yaklaşımda kullanıcıdan sadece bir parametreyi belirlemesi beklenmektedir.Geliştirilen yöntem CluStream, DenStream ve ClusTree algoritmaları ile karşılaştırıldığında daha iyi sonuç verdiği tespit edilmiştir.DBSTREAM veriyi mikro-kümelere bölerken sadece mikro-kümeler arası mesafeyi almaz aynı zamanda orijinal veriler arasındaki yoğunluk bağlantılarını da alır.Veriler w boyutunda parçalara ayrılır.Yani geriye dönüş w süresince olabilir.Burada küme birleştirme ve ayırma (merge ve split) k-means ile yapılır.Her veri gelişinde graf yapısı kontrol edilmekte ve eğer gerek varsa bu yapı güncellenmektedir.Offline aşamada ise kullanıcıya mikro-kümelerden oluşan makro kümeler sunulur.LLDStream boyut indirgemeyi LDA (Linear Discriminant Analysis)'dan faydalanarak yapar.Bunun için akan veriden birbiri ile alakalı nitelikleri bulmak için frekans spektrumunu bulur.Online bölüm kayan pencere ile akan verilerin spektral bileşenlerini hesaplar.SPE-Cluster dinamik yapısı ile kümeleme adaptasyonu (concept evolution) ve küme sayısı problemlerine çözüm üretmektedir.Jurgen ve ark.Veriyi özetlemek için veriler arasındaki uzaklığı ve Discrete Fourier Transform (DFT) katsayısını kaydeder.Üstten aşağıya (top-down) bir yapıyı destekler, hiyerarşik bir ağaç yapısındadır.Hiyerarşik ağaçta her bir kök (node - burada küme) uzaklığa bakarak ayırır.Var olan kümelerin çapına bakarak kümelerin birleştirilmesine veya ikiye ayrılmasına karar verir.Bunun için ortalama, standart sapma ve her küme için nokta sayısına bakar.Benzerlik ölçütü olarak ağırlıklandırılmış korelasyon kullanır.Belirli bir durum oluştuğunda kümeleri böler veya birleştirir.Çok sayıda parametre var ve parametrelerin çok doğru seçilmesi gerekir.Özellikle banka, sağlık kuruluşları, güvenlik ve sağlık kurumları gibi kuruluşlar için veriyi akarken anlamlandırmak gerekir.Çünkü bu tür kuruluşlarda zamana karşı bir yarış söz konudur.Bunun yanında bilim, ticaret, meteoroloji, teknoloji ve kamu yönetimi gibi pek çok alanda yaygın olarak kullanılmakta ve kullanım alanı gün geçtikçe artmaktadır.Literatürdeki çalışmaların çoğu online-offline olarak iki bileşenden oluşmaktadır.Tamamen online çalışan yöntemler de önerilmektedir.Ancak bu tür yöntemler ya istenen başarıyı vermemektedir ya da performansı arttırmak için özetleme yöntemleri kullanıldığından verinin önemli bir kısmı yok sayılmaktadır.Kısaca verinin nasıl özetleneceği de yine çalışmaya açık alanlardan birisidir.Geliştirilen yöntemlerde kullanılan parametrelerin seçimi ve atanması da yine güncel problemlerden birisidir.Çoğu çalışmada çok sayıda parametrenin belirlenmesi gerekmektedir.Bu parametrelerin doğru belirlenmemesi başarıyı doğrudan etkilemektedir.Parametreleri azaltmak ve bu parametrelerin mümkün olduğunca model tarafından belirlenmesi de önemli problemlerden birisidir.Ayrıca sapan verileri tespit etmek ve performansı düşürmeden çok boyutluluğu desteklemek de güncel problemler arasındadır.

SUMMARY 2: "Blok Zinciri Teknolojisi: Kullanım Alanları, Açık Noktaları ve Gelecek Beklentileri", (DOI: 10.31590/ejosat.423676)

New York Times, The Economist ve Wired gibi birçok ünlü haber dergisinde Bitcoin'in bu gizemli öyküsü çeşitli defalar yayınlanmış ve Bitcoin'in arkasındaki kişi veya kişilerin kimler olduğuna dair çeşitli tahminler yapılmıştır.Günümüzde, öncelikle merkez bankaları olmak üzere çeşitli devlet kurumları, birçok büyük banka ve teknoloji şirketi blok zinciri teknolojisine yatırım yapmakta ve bu amaçla çeşitli işbirlikleri oluşturmaktadır.Bu gelişmeler ve blok zinciri teknolojisinin kullanım alanları makalenin ilerleyen bölümlerinde detaylı şekilde tartışılacaktır.Bu çalışmaların önemli bir kısmı da, dünya yazınındaki yayınlarla paralel bir şekilde, ağırlıklı olarak Bitcoin'in hukuki ve ekonomik durumuna odaklanmaktadır.Her yeni teknolojide olduğu gibi, blok zinciri teknolojisinin sağlıklı şekilde gelişebilmesi ve yaygınlaşabilmesi de ancak akademik alanda yapılacak çalışmalarla desteklenmesiyle mümkündür.Bu nedenle, blok zinciri teknolojisini sistematik olarak inceleyen, araştırma alanlarını, açık konuları ve kullanım alanlarını tartışan akademik çalışmalara ihtiyaç duyulduğu gözlemlenmektedir.Standart tek bağlı liste yapısında, listenin her elemanı, kendinden sonra gelen elemanı bir işaretçi yordamıyla işaret eder.Bu şekilde listenin başlangıç elemanından kuyruk elemanına kadar bütün elemanlar birbirlerine bağlanmış şekildedirler. Blok zinciri yapısında ise her eleman (blok), sadece sonraki bloğu işaret etmez, aynı zamanda o bloğun özet (hash) değerini de saklar.Özet-işaretçi yapısı, bu tür bir değişikliğe izin vermez, daha doğrusu bu tür bir değişiklik yapıldığı kolaylıkla anlaşılır.Çünkü yeni eklenen bloğun özet değeri, bu yeni bloğu işaret eden özet-işaretçisinin işaret ettiği değerden farklı olacaktır.Bu özellik blok zincirinin güvenli bir yapı olmasını sağlayan önemli etmenlerin başında gelmektedir.Bu bilgiler daha sonra açık muhasebe defterine kaydedilir.Bu şekilde, tüm finansal hareketler açık muhasebe defterine kaydedilmiş ve merkezi bir otorite olmaksızın finansal bir işleyiş kurulmuş olur.Burada en önemli nokta, açık muhasebe defterine kaydedilen işlemlerin doğru işlemler olması, kötü niyetli eşler tarafından üretilen sahte veya hatalı işlemler olmamasıdır.Aksi halde, sistem finansal hırsızlık olaylarına açık hale gelir.Diğer tüm eşler, kendilerine gelen işlemleri değerlendirir ve doğru ise kabul ederek kendilerinin üreteceği yeni işlem bloklarına eklerler.Bu finansal teşvik yardımıyla dağıtık uzlaşı sisteminin, doğru işlemleri paylaşan güvenilir eşleri destekleyerek daha güvenli hale gelmesi hedeflenmiştir.İşlem ücretleri opsiyoneldir ancak blok ödülleri belli periyotlarla yarılanarak devam etmektedir.Gerçek dünyada yapılan işlemde güvenlik önlemleri, fiziksel paranın güvenli şekilde imal edilmesine ve kolay şekilde taklit edilememesine odaklanmıştır.Yani A kişisi B kişisine para transferi yaparken, işlemi kendi dijital imzasıyla imzalayarak güvenli hale getirir.Dijital imzalama işlemi, yapılan işlemlerin güvenliğini sağlarken; ayrıca işlem sahiplerinin de gizliliğini sağlamaktadır.Bitcoin ile yapılan tüm işlemler açık muhasebe defterine kaydedilir ve adından da anlaşılacağı üzere bu deftere isteyen herkes erişebilir.Ancak işlemlerde gerçek kişi bilgileri veya bundan türetilen bir bilgi yerine, dijital imzalar yer aldığı için finansal hareketlerin gerçekte kimler tarafından yapıldığı bilinmez.Teşviği hangi eşin alabileceği konusu ise oldukça önemli bir problemdir.Dolayısıyla teşvik edilecek eşlerin seçimi, rastlantısal şekilde olabilmeli ve sistemde hiçbir eş ya da eş grubu tekel olarak davranamamalıdır.Eşlerin, sisteme yeni blok ekleyebilmeleri, çözümü için yüksek işlemci gücüne ihtiyaç duyacak karmaşık bir özet-bulmacanın çözülmesiyle mümkün olmaktadır.Yüksek işlemci gücüne dayalı bir problemin çözümü, teşvik kazanmak isteyen kimselerin güçlü özelliklerdeki çok sayıda bilgisayarı Bitcoin sistemine dâhil etmelerini sağlamıştır.Ancak sürekli yeni eşlerin katılarak, sistemi büyüttüğü bir dağıtık sistemde, belirli eş veya eşlerin tekel olabilmeleri çok kolay olmamaktadır.Emek ispatı hali hazırda en popüler blok zinciri platformlarında kullanılan blok üretim ve doğrulama mekanizması olmakla birlikte, yüksek enerji tüketimine neden olmakta ve özel donanım gereksinimleri ortaya çıkarabilmektedir.Bu durum ise çeşitli eleştirilere neden olmakta ve alternatif çözümler önerilmektedir.Alternatif bir yöntem olarak öne çıkan hisse ispatı (proof of stake – PoS), bloğu üreten eşin ilgili blok zinciri ağı üzerinde sahip olduğu pay ile orantılı olarak geçerlilik onay yetkisi vermektedir.Bu yöntemde, yüksek pay sahibi eşlere sürekli bir avantaj sağlanmasının önüne geçmek için akış içerisindeki hesaplamalarda kullanılmak üzere yaş (age) kavramı da geliştirilmiştir.Bu sayede, herhangi bir blok üretimi için kullanılan pay kapsamındaki kripto paraların yaş değerleri sıfırlanır ve ancak belirli bir süre sonra tekrar yaş değeri kazanmaya başlarlar.Bu durum alışık olduğumuz hesap yönetiminden biraz daha farklıdır.Örneğin banka hesabınızda bir bakiyeniz vardır; hesabınıza gelen paralar bakiyenizi arttırırken, hesaptan çıkan paralar da bakiyenizi azaltır.Bir para harcaması yapacağınız zaman, eğer harcamak istediğiniz tutar bakiyenizden büyükse işlemi yapamazsınız.Bitcoin işlemleri ise bir bakiye yönetimi gibi çalışmaz, her işlem kendi içinde girdi ve çıktılarını belirtir.Metadata bölümünde, işlemin tekil numarası, girdi sayısı ve çıktı sayısı gibi işleme ait temel veriler yer almaktadır.Anahtar yönetimiyse bilişim güvenliği alanında başlı başına önemli bir konudur.Zira anahtarların başkaları tarafından ele geçirilmesi önemli riskler içerir.İnternette ücretsiz olarak birçok Bitcoin cüzdanı yazılımına erişilebilir.Bu alanda yenilikçi bir çözüm de sadece Bitcoin'leri değil, paralelinde birçok farklı kripto parayı yönetmeyi sağlayan özelliği ile öne çıkan Exodus (http://www.exodus.io) isimli cüzdandır.Bunun yanısıra Electrum (https://electrum.org), Jaxx (https://jaxx.io/) ve Mycelium (https://wallet.mycelium.com/) gibi cüzdanlar da popüler cüzdan yazılımlarıdır.Bu nedenle sıcak ve soğuk cüzdan kavramları geliştirilmiştir.Soğuk cüzdanlar ise daha güvenli olması amacıyla, sürekli çevrimiçi olmayan ve sadece gerektiğinde sıcak cüzdanlarla para transferi yapmak için çevrimiçi olan cüzdanları tarif eder.Trezor (https://trezor.io/) ve Ledger (https://www.ledgerwallet.com/) günümüzde yaygın olarak donanım tabanlı soğuk cüzdanlardır.Kısmi-

merkezi bok zinciri yapılarında, diğer bir adıyla konsorsiyum blok zincirleri, dağıtık uzlaşı yöntemi yerine sadece önceden belirlenmiş sınırlı sayıda eşin, uzlaşı sistemini yönettiği yapılardır.Bu tür yapılarda blok zinciri verisi herkese açık olabileceği gibi verilerin erişilebilirliğinin de çeşitli şekillerde kısıtlandığı karma blok zinciri yapıları oluşturulabilir.Verileri okuma hakkı ise herkese açık olabileceği gibi çeşitli şekillerde kısıtlanabilir.Bu platformlar, açık kaynaklı olup olmamaları, fiyatlandırma yapısı, destekledikleri programları dilleri ve destekleri blok zinciri yapılarına (açık, hibrid, özel) göre farklılaşmaktadır.Ethereum ve Hyperledger, bu alanda en yaygın kullanılan, en bilindik alternatiflerdir.Kripto paralar dışında blok zinciri teknolojisinin özellikle güven sorunu yaşanan ve bu sorunu aşmak için aracı kişi ve kurumların (intermediaries) yer aldığı birçok iş alanında kullanılabileceği öngörülmektedir.Blok zinciri teknolojisini kullanan birçok kavram kanıtlama ve prototip ürün çalışmaları yapılmasına rağmen gerçek hayatta kullanılan uygulamalar henüz hayata geçmemiştir.Dolayısıyla blok zinciri teknoloji kullanan iş fikirleri oldukça ilgi çekmesine rağmen, bu fikirlerin hayata geçmesi için yatırım bütçelerine ve zamana ihtiyaç vardır.Uluslararası ticaretin finansmanı konusunda ise çeşitli bankaların bir araya gelip konsorsiyumlar oluşturduğu duyurulmaktadır.Bitcoin Betik diliyle tanımlanabilen akıllı kontrat yapısı sayesinde, kontrat içeriğinde belirtilen gerekli mantıksal koşulların sağlanması durumunda, amaçlanan işlemin hayata geçmesi sağlanmaktadır.Merkezi bir otoriteye gerek duymaksızın, farklı paydaşlar arasında dijital sözleşmelerin tanımlanabilmesi ve sözleşme koşullarının takip edilerek sonucuna göre hedeflenen aksiyonların otomatik olarak hayata geçirilebilmesi, blok zinciri teknolojisinin en çok heyecan yaratan uygulama alanları arasındadır.Diğer bir taraftan, blok zinciri teknolojisinin etkilerinin mevcut ürün ve hizmetleri iyileştirmenin çok daha ötesinde olacağını savunan ve buna yönelik düşünceler geliştirilen kişi ve kurumlar da vardır.IOTA, komisyon ücretleri olmadan, dağıtık ve ölçeklenebilir bir altyapı sunmakta ve nesnelerin açık muhasebe defteri (ledger of things) olarak adlandırmaktadır.Bu gelişmelere biraz daha mesafeli duran ve blok zinciri teknolojisi etrafında oluşan yüksek beklentilerini sorgulayan ve blok zinciri teknolojisinin net faydalarının görülebileceği projelere odaklanılmasını öneren görüşler de vardır.Danışmanlık şirketlerinin raporlarının yanı sıra, yeni bir teknolojinin gelişimini destekleyen en önemli konulardan biri de standardizasyon çalışmalarıdır.Tüm bu gelişmeler, dünya çapındaki önemli kuruluşların blok zinciri teknolojine önem verdiklerini, yaygınlaşmasını desteklediklerini, bu teknolojinin geleceğine dair olumlu beklentileri olduğunu ve bu yönde önemli yatırımlar yaptıklarını göstermektedir.Sıfır Bilgi Kanıtı algoritması, özetle bildiğiniz bilgiyi bir başkasına, bilgiyi ona vermeden ispat etmek olarak tanımlanabilir.Sıfır Bilgi Kanıtı algoritması, erişilebilir blok zinciri verileri üzerinden işlem ve grafik analizi gibi yöntemlerle çeşitli çıkarımlar yapmanın önüne geçmede ve blok zinciri işlemlerinin mahremiyeti artırmada etkili bir algoritmadır.Sistemin güvenli ve dağıtık bir şekilde çalışabilmesi için bütün eşlerde aynı özellik ve algoritmaları çalıştıran açık kaynaklı yazılımlar çalışmalıdır.Çeşitli nedenlerle bu yazılımın özellikleri, parametreleri veya kullandığı algoritmalarda değişiklikler veya geliştirmeler yapılması gerekebilmektedir.Bu durumlarda yapılan değişikliklerin bazen yumuşak (soft-fork) bazen sert (hard-fork) şekilde yapılabilir: Diğer bir ifadeyle yapılan değişiklik daha önce geçerli olan bir bloğu/işlemi geçersiz hale getirebilir (soft-fork) veya daha önce geçersiz olan bir bloğu/işlemi geçerli hale getirebilir (hard-fork).Bu krizler sisteme olan güveni azaltabilmekte ve blok zinciri teknolojisine olan desteğin de azalmasına neden olabilmektedir.Performans ve Ölçeklenebilirlik (Performance and Scalibility) Bitcoin, her ne kadar gittikçe artan bir kullanım oranına sahip olsa da, dünyadaki tüm finansal işlemlerin hacmi düşünüldüğünde henüz düşük diyebileceğimiz seviyelerde finansal hacime ve sınırlı işlem sayılarına sahip bir para birimidir.Ölçeğin artması böylesine büyük dağıtık bir sistem üzerinde koşan algoritmaları saniyede binlerle ifade edebilecek işlem seviyeleri ile karşı karşıya bırakacaktır.Pratik hayatta da blok zincirini iyileştirmeye yönelik örnekler ortaya çıkmaya başlamıştır.Bu teknoloji hızlı ve ucuz işlem ücretleriyle işlem yapmak isteyen eşlerin, kendi aralarında ve blok zincirine ağına bağımlı ödeme kanalları açmaları ve işlemlerini bu ödeme kanalları üzerinde yapmaları esasına dayanır.Ödeme kanalları üzerinde işlem güvenliğini sağlamak için çoklu imzalama ve iki yönü ödeme yapılabilmesi gibi mekanizmalar da tasarlanmıştır. Bitcoin ve kripto para kavramları, son dönemde ekonomi ve finans dünyası için en ilgi çekici konuların arasında yer almaktadır.Bitcoin'e hayat veren ve açık muhasebe sistemi kavramını hayatımıza sokan blok zinciri teknolojisi ise, gerek dağıtık mimarisi gerekse güvenlik özellikleriyle büyük mühendislik firmaları ve danışmanlık şirketlerinin ilgi odağı olmuştur.On yıl gibi kısa bir sürede birçok ürün ve hizmetin altyapısını oluşturması beklenen blok zinciri teknolojisi, mühendislik anlamında önemli gelişmelerin yaşanacağı, büyük dönüşümlerin gözlemleneceği bir alan olarak öne çıkmaktadır.Dünyanın önde gelen üniversitelerinde bu alanda derslerin açılmaya başlandığını ve araştırma laboratuvarlarının kurulduğunu gözlemlemek mümkündür.Ülkemizde ise sınırlı sayıda olmakla birlikte olumlu gelişmeler yaşandığını söyleyebiliriz.) de olduğu bilinmektedir.Blok zinciri teknolojisi, yapay zekâ ve nesnelerin interneti teknolojileri gibi çok büyük değişimlere yol açabilecek teknolojiler arasında gösterilmektedir.Bu alandaki çalışmaların teşvik edilerek yaygınlaştırılması, dünyada yaşanan önemli bir teknolojik dönüşüm sürecinde ülkemizin hak ettiği yeri almasını sağlayacaktır.

SUMMARY 3: "Bluetooth Düşük Enerji Teknolojisine Sahip İşaretçi ve Akıllı Telefon Temelli Öğrenci Yoklama Sistemi**", (**DOI: 10.17671/btd.94742**)**

Öğrencilerin derse devam takip işlemleri çoğunlukla yoklama listelerine imza atma yöntemi ile yapılmaktadır.Öğrenci sayısının çok olduğu büyük sınıflarda yoklama işlemi oldukça zaman almaktadır.iBeacon cihazlar konumlandırıldıkları ortamda sürekli Bluetooth paketi yayarlar ve bu paketi alan akıllı telefon gibi mobil cihazlardaki uygulamaların tetiklenerek çalışmasını sağlarlar.Bu nedenle işaretçiler BLE beacon olarak ta adlandırılmaktadır.Geliştirilen sistem devam takip işlemlerinin otomatik olarak gerçekleştirilmesi, öğrencinin katılım süresinin saat bazında tam olarak belirlenmesi, yoklama işlemi için harcanan sürenin kısaltılması gibi birçok avantaj sunmaktadır.Bununla birlikte kullanıcıların BLE teknolojisini destekleyen mobil cihazlara sahip olma zorunluluğu sistemin bir dezavantajı olarak göze çarpmaktadır.Makalenin organizasyonu şu şekildedir.Son bölümde ise sonuçlar ve değerlendirmeler verilmektedir.Mobil cihazlardaki uygulamalar, işaretçi cihazlardan alınan kablosuz sinyaller ile tetiklenir/çalıştırılır.Tüm ağı tanımlamak için kullanılır ve genellikle üretici bilgisini içerir.İlk olarak kullanıcılar, mobil uygulamaya kullanıcı adı ve şifre ile giriş yapmalıdırlar.Veritabanı olarak MySQL veritabanı kullanılmıştır.Kullanıcı adı ve şifre ile mobil uygulamaya giriş yapıldıktan sonra ilk olarak kullanıcı bilgileri gelmektedir.Bu durum kullanıcıya şekilde de görüldüğü üzere bildirim mesajları ile aktarılmaktadır.Bu yazılımının tasarım ve geliştirilmesinde bootstrap, php, html teknolojileri kullanılmıştır.Bu arayüzde ilgili dersi alan öğrenciler ve öğrencilerin toplam katılım yüzde oranları görülmektedir.İşaretçiler reklam, lojistik, ürün, tarihi eser ve yerlerin tanıtımı, navigasyon gibi çok geniş uygulama alanı sunmaktadır.Bu çalışmada, kullanımı hızla yaygınlaşan ve güncel bir teknoloji olan işaretçilere dayalı olarak geliştirilen yoklama ve katılım sistemi sunulmuştur.Böylelikle öğrencinin hangi saat sınıfta olduğu hangi saat ayrıldığı izlenebilmektedir.Geliştirilen sistemin en büyük dezavantajı ise kullanıcıların BLE teknolojisine sahip mobil cihaz kullanımıdır.Günümüzde yeni cihazların çoğunda BLE teknolojisinin standart olarak sunulması bu dezavantajın hızla giderileceğini göstermektedir.