



APPLICATION OF TEXT SUMMARIZATION METHODS TO ARTICLES IN THE FIELD OF AGRICULTURE

Ebru TEMİZZHAN^{1*}, Mehmet MENDEŞ¹


¹Canakkale Onsekiz Mart University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetics Unit, 17100, Çanakkale, Türkiye


Abstract: This study evaluates the applicability of text summarization algorithms to articles in the field of agriculture. The abstract and conclusion sections of articles on the topics of "agriculture" and "organic agriculture" were analyzed using extractive text summarization algorithms: TextRank, LexRank, Luhn, and LSA. The summaries generated by each algorithm were compared using the cosine similarity measure. These similarities were then visualized on a 2-dimensional plane using a Venn diagram. The findings indicate that there are similar tendencies in the algorithms' selection of content-focused sentences for both agriculture and organic agriculture articles. Notably, it was observed that among the text summarization algorithms, LexRank and LSA produced more consistent results across both datasets. In conclusion, it has been demonstrated that summarization methods can be effectively applied in agricultural research to reduce information density.

Keywords: Agriculture, Extractive text summarization, TextRank, LSA, Cosine similarity

*Corresponding author: Canakkale Onsekiz Mart University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetics Unit, 17100, Çanakkale, Türkiye

E mail: ebrutemizhan@gmail.com (E. TEMİZZHAN)

Ebru TEMİZZHAN  <https://orcid.org/0000-0001-7068-0492>

Mehmet MENDEŞ  <https://orcid.org/0000-0003-0109-9914>

Received: January 29, 2026

Accepted: May 07, 2026

Published: May 15, 2026

Cite as: Temizhan, E., & Mendes, M. (2026). Application of text summarization methods to articles in the field of agriculture. *Black Sea Journal of Engineering and Science*, 9(3), 1369-1377.

1. Introduction

Nowadays, the rapid access to information and the exponential increase in data volume have created a need for more effective examination of academic studies. According to 2023 statistics, approximately 12,600 academic articles are published daily in reputable international databases worldwide, of which around 5% (~600 articles) are produced in the field of agriculture. In disciplines with intensive interdisciplinary research, such as medicine, engineering, veterinary science, and agriculture, there is a growing need for advanced information processing tools to facilitate literature tracking and enable rapid comprehension of lengthy articles. Automatic text summarization techniques have gained importance in providing researchers with concise and more meaningful representations of relevant articles, facilitating their investigations. Automatic text summarization techniques aim to generate concise, dense, and readable representations while preserving the main idea of the text. Early studies on summarization have shown that short summaries can nearly double decision-making speed (Mani, 2001; Mani et al., 2002). Recent research, on the other hand, has demonstrated that modern summarization models can reduce human bias and provide high scalability across large and diverse datasets (Zhang et al., 2020; Krishna et al., 2021; Luo et al., 2024). Extractive summarization methods such as TextRank, LexRank, Luhn, and LSA are among the most

frequently applied approaches for summarizing scientific texts (Luhn, 1958; Landauer et al., 1998; Erkan and Radev, 2004; Mihalcea and Tarau, 2004). TextRank, with its graph-based structure, calculates relationships between sentences and identifies the most central statements in the text, producing particularly stable results in long technical documents (Mihalcea and Tarau, 2004; Wan and Yang, 2008). Similarly, LexRank employs a graph structure combined with a voting system that measures sentence similarity; recent studies have shown that this method can balance topic distribution effectively in multi-author academic texts (Erkan and Radev, 2004; Qazvinian and Radev, 2008). The Luhn algorithm, which is based on word frequencies and density zones, remains a classical method, but modern adaptations have enhanced its ability to accurately identify key sentences, especially in scientific texts containing dense technical terminology (Luhn, 1958; Rao et al., 2022). Latent Semantic Analysis (LSA) maps the conceptual structure of a text into a mathematical space, revealing abstract semantic relationships. Recent studies indicate that LSA can consistently resolve conceptual clusters in interdisciplinary articles (Landauer et al., 1998; Steinberger and Jezek, 2004; Allahyari et al., 2017). Considering the structural characteristics of agricultural literature, including dense terminology, interdisciplinary content, and methodological diversity, it is necessary to systematically evaluate the performance of these



algorithms in a domain-specific context (Allahyari et al., 2017).

In this study, four extractive text summarization algorithms (TextRank, LexRank, Luhn and LSA) were applied to academic articles related to agriculture and organic farming. The similarities between the summaries generated by these algorithms were analyzed using the cosine similarity metric to examine which method produced more consistent results. The primary objective of this study is to identify the most suitable summarization method or methods that facilitate a more efficient examination of academic texts in the field of agriculture.

2. Materials and Methods

The material for this study consists of Turkish articles collected from the Google Scholar database using the keywords "agriculture" and "organic farming" (Temizhan, 2024). Information regarding the dataset is given in Table 1. In the statistical analysis of this study, the hardware used was an Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz 2.59 GHz processor, 16 GB RAM, a 64-bit operating system, and an NVIDIA GeForce RTX 2060 graphics card. The software used was Python (3.12.1) and Jupyter Notebook (6.0.3) (Jupyter Notebook 2024).

Table 1. Article datasets

Keywords	Year	N
"Tarım"	2011-2020	496
"Organik Tarım"	2010-2020	522

The flowchart of the methodology is presented in Figure 1. The datasets obtained during the data collection stage are listed in Table 1. In the data extraction stage, only the "abstract" and "conclusion" variables were selected from these datasets to construct the datasets intended for analysis. Subsequently, several text preprocessing steps were applied to this data. Within this scope, special tweet markers (@, #), emojis, and URL links (https) were removed from the texts to clean them. Following this, the abstract and conclusion sections of the articles were summarized separately using four selected automatic text summarization algorithms. The semantic similarity between the generated summary outputs was evaluated using the cosine similarity metric as a quantitative measure. In the final stage, the cosine similarity values between the summaries were visualized on a two-dimensional plane, and the contextual proximity of the algorithm outputs was illustrated using a Venn diagram. Following these visual analyses, the results were interpreted to derive meaningful insights from the summary texts.

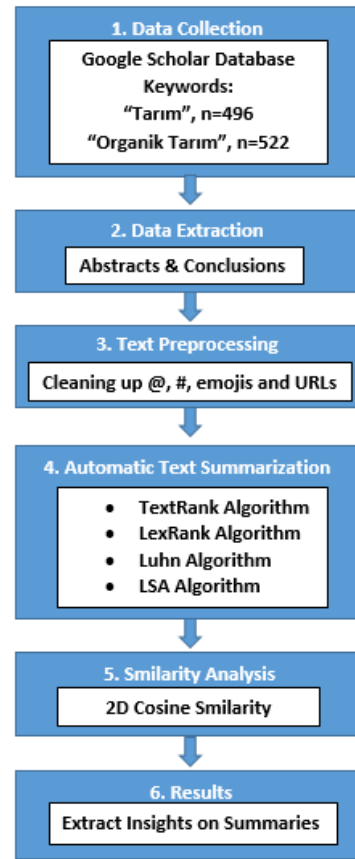


Figure 1. Methodology flowchart.

2.1. Text Summarization Algorithms

2.1.1. TextRank algorithm

TextRank is a graph-based algorithm. It works by converting text into a weighted word or sentence graph. It evaluates the similarities between sentences within the text and selects the most important sentences. It is used in tasks such as text summarization, document sorting, and keyword extraction. The working principle of the TextRank algorithm is based on the mathematical formula given in equation 1 (Mihalcea and Tarau, 2004).

$S(V_i)$: i. score of sentence,
 d : damping factor (usually 0.85),
 $In(V_i)$: set of sentences referring to sentence i.,
 $Out(V_j)$: set of sentences referring to sentence j.,

$$S(V_i) = (1 - d) + d \sum_{V_j \in In(V_i)} \frac{S(V_j)}{|Out(V_j)|} \quad (1)$$

2.1.2. LexRank algorithm

The LexRank algorithm identifies significant sentences in texts by measuring similarities between them. Sentences are represented by a similarity matrix. It is used in operations such as text summarization, document sorting, and document similarity (Erkan and Radev, 2004). The working principle of the LexRank algorithm is based on the mathematical formula given in equation 2.

$sim(C_i, C_j)$: the similarity between sentences i. and j.,

w_k^i : the weight of the k. word in sentence i.,

w_k^j : the weight of the k. word in sentence j.,

$$sim(C_i, C_j) = \frac{\sum_k w_k^i \cdot w_k^j}{\sqrt{\sum_k (w_k^i)^2} \cdot \sqrt{\sum_k (w_k^j)^2}} \quad (2)$$

2.1.3. Luhn algorithm

Luhn's algorithm selects prominent sentences based on frequently occurring words. It is used in operations such as keyword extraction and document sorting (Luhn, 1958). The working principle of Luhn's algorithm is based on the mathematical formula given in Equation 3.

$Score(C_i)$: the score of the i . sentence,,

$freq(k)$: the frequency of word k ,

$$Score(C_i) = \sum_{k \in C_i} freq(k) \quad (3)$$

2.1.4. LSA (Latent semantic analysis) algorithm

The LSA (Latent Semantic Analysis) algorithm evaluates semantic similarities between words and documents. It provides text summarization by applying Singular Value Decomposition (SVD) to the Tf-Idf (Term frequency-Inverse document frequency) matrix. It is used in processes such as document similarity, summarization, and information extraction (Deerwester et al., 1990). The working principle of the LSA algorithm is based on the mathematical formula given in equation 4.

A : Word-document matrix,

U : Left singular vectors matrix,

Σ : Diagonal matrix (single values),

V : Right singular vectors matrix,

$$A = U \Sigma V^T \quad (4)$$

2.2. Similarity Measurement and Visualization

The similarity between text summaries was measured using the cosine similarity metric. Cosine similarity is based on calculating the cosine of the angle between two vectors represented in n -dimensional space. Widely used in the comparison of text documents, cosine similarity enables the quantitative measurement of similarity between documents (Rios et al., 2010). For two vectors A and B , cosine similarity is calculated using the formula in equation 5.

$$cosine\ similarity(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}} \quad (5)$$

The cosine similarity metric takes a value between 0 and 1. As the similarity measure approaches 1, it indicates a high degree of content similarity between the two texts. These measured similarity ratios have been expressed as percentages (%). As a result of the visualization analysis, it has been observed that texts represented by points located close to each other in the two-dimensional plane show a greater degree of similarity.

3. Results

Automated summary outputs for the agricultural dataset are given in Table 2, automated summary outputs for the organic farming dataset are given in Table 3, cosine similarity ratios are given in Table 4 and the two-dimensional representation of cosine similarity ratios is given in Table 5.

Table 2 shows the outputs obtained using automated text summarization methods for the summary variable of the agricultural dataset, range from [283-309] characters. The number of words for the summary to be generated using the TextRank algorithm was set to 300. The resulting summary output consisted of 5 sentences and 283 characters. The number of sentences for the summary to be generated by the LexRank algorithm was set to 40. The resulting summary output consisted of 11 sentences and 267 characters. The number of sentences for the summary to be generated by the LSA algorithm was set to 8. The resulting summary output consisted of 8 sentences and 285 characters. The number of sentences for the summary to be generated by the Luhn algorithm was set to 3. The resulting summary output consisted of 3 sentences and 309 characters.

When the cosine similarities of the outputs for the summary variable of the agricultural dataset shown in Table 4 were examined, it was observed that the TextRank algorithm exhibited a high degree of similarity with the other algorithms. The similarity rates reached 97.81% between the TextRank and LSA algorithms, 94.68% between the TextRank and Luhn algorithms and 94.18% between the TextRank and LexRank algorithms.

In Table 2, the outputs obtained using automatic text summarization methods for the conclusion variable of the agricultural dataset range from [198-266] characters. The number of words for the summary to be generated using the TextRank algorithm was set to 250. The resulting summary output consisted of 5 sentences and 266 characters. For the LexRank algorithm, the number of sentences to be generated in the summary was set to 100. The resulting summary output consisted of 11 sentences and 198 characters. For the LSA algorithm, the number of sentences to be generated in the summary was set to 6. The resulting summary output consisted of 6 sentences and 253 characters. For the Luhn algorithm, the number of sentences to be generated in the summary was set to 1. The resulting summary output consisted of 1 sentence and 253 characters.

When the cosine similarities of the outputs for the conclusion variable of the agricultural dataset in Table 4 were examined, high similarity rates were observed between the LSA and Luhn algorithms 97.99%, between the TextRank and LSA algorithms 97.50% and between the TextRank and Luhn algorithms 97.01%.

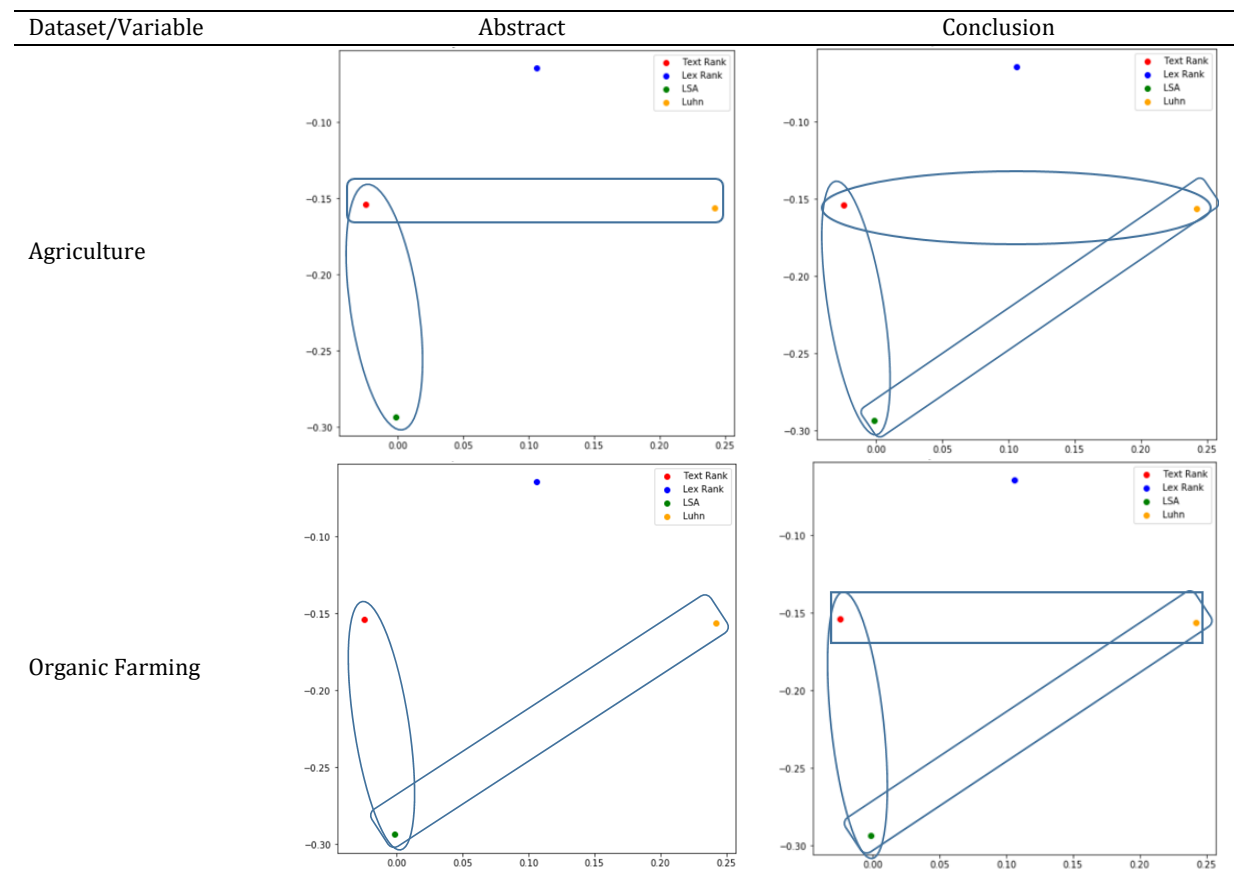
Table 3. Automatic summarization for organic farming dataset

Variables	Algorithms	Automatic Summarizations
		<p>Sahip olduğu tarım arazilerini ve su kaynaklarını kirletmeden ve doğal kaynaklarını tahrip etmeden gelecek kuşaklara aktarmak, sürdürülebilir tarımsal kalkınmayı sağlamak, topluma sağlıklı gıda arz etmek ve katma değeri yüksek tarımsal üretim yapılarak üretticilerin ekonomik gelir düzeyini yükseltmek amacıyla organik tarımın geliştirilmesi ve yaygınlaştırılması her ülke için büyük kazanımlar oluşturmaktadır.</p> <p>Organik parsellerde bitki koruma materyali olarak Organik Tarım Kanununda izin verilen preparatlar, organik parçacıklar ve parçacıklar 0 ürtünde rüncatlı olan sentetik preparatlar kullanılmaktadır. 9 yılın toprak analiz sonuçları değerlendirildiğinde su ile donmuş k, tut ve kireç değerlerinde her iki sisteme herhangi bir değişiklik olmamıştır.</p> <p>Bu çalışmada, insan beslenmesi ve sağlığı açısından çok önemli bir ürün olan zeytinin yetiştirilmesinde, konvansiyonel yöntemle re alternatif olarak organik tarım yöntemlerinin uygulanabilirliğinin belirlenmesi amaçlanmıştır.</p> <p>Organik Tarım genel olarak mevcut olan ekolojik dengeyi korumak için tasame ve çevreyi tehdit eden türden sanayileşmenin en o olduğu bölg elerden biridir.Bununla birlikte, Doğu Karadeniz Bölgesi farklı coğrafik yapıya, korumuş zengin Flora ve Faunası yanında kültü r bitkisi çeşitliliği ile organik tarım açısından önemli bir potansiyele sahiptir.</p> <p>Türkiye’de organik hayvancılık açısından büyük bir potansiyele bulunmaktadır.</p> <p>Türkiye organik tarımında çok büyük bir potansiyele sahiptir.</p> <p>Toprak analiz sonuçlarına göre, organik parsellerde sertifikalı gübre ve yeşil gübre uygulanmıştır.</p> <p>Endüstriyel veya konvansiyonel tarım, günümüzde en yaygın olarak uygulanan tarım yöntemidir.</p> <p>Çalışmada, 5 farklı organik bitki besleme uygulaması A (organik gübre + çiftlik gübresi), B (organik gübre + yeşil gübre), C (organik gübre + humi k asit), C (organik gübre + çiftlik gübresi), D (organik gübre + yeşil gübre) ve E (organik gübre + organik gübre + yeşil gübre) ile konvansiyonel yetiştiricilik verim ve potansiyel özellikler açısından karşılaştırılmıştır.</p> <p>Bu çalışmanın ana amacı, Türkiye’de organik tarım ve organik bitkisel ürünlerin geliştirilmesi için araştırılmasıdır.</p> <p>Türkiye ile ilgili olarak 2002-2009 döneminde organik tarımın gelişimi ele alınmıştır.</p> <p>Dünyada ve Avrupa’da organik tarım içerisinde organik hayvansal üretimin payı, her geçen gün artmaktadır.</p> <p>Gerek insan sağlığı gerekse bitki ve çevre sağlığına olumlu katkılarının dolaylı olarak organik tarım alanı ilgi son yıllarda giderek artmaktadır.</p> <p>Bu ürünlerin başında da organik çay üretimi gelmektedir.</p> <p>Bu derlemede yeşil gübreleme hakkında bilgi verilmektedir.</p> <p>Organik gübrelere olan talep son yıllarda giderek artmaktadır.</p> <p>kullanılmaktadır.</p> <p>Son yıllarda bu amaçla organik herbisitler kullanılmaktadır.</p> <p>Ülkemizde organik ürünler 1984-1985 yıllarında başlamıştır.</p> <p>konvansiyonel tarımın yanı sıra, organik tarım yapan işletme sayısı bölgede artış göstermektedir.</p> <p>Bu çalışma, serada organik domates yetiştiriciliğinde farklı bitki yoğunluklarına (1,07, 1.14, 2.38, 2.05, 3.09, 2.37, 1.90 ve 2.42 bitki m²) bazı toprak özellikleri ve verim üzerine etkilerinin belirlenmesi amacıyla yürütülmüştür.</p> <p>Çalışma 2005-2007 yılları arasında yürütülmüştür.</p> <p>L. Amaç: Ülkemizde organik gıda talebi son yıllarda giderek artmaktadır.</p> <p>Bu çalışma organik kaba yem üretimi amacıyla yetiştirilen bazı fiğ türlerinin yem verimi ve kalitesi üzerine farklı organik gübre kaynaklarının etkisini belirlemek amacıyla yürütülmüştür.</p> <p>kontrol, 2. olarak belirlenmiştir.</p> <p>Toplam 138 üreticisi ile anket yapılmıştır.</p> <p>Organik ürünleri olan talep son yıllarda giderek artmaktadır.</p> <p>Bu nedenle konvansiyonel tarım alternatif bir yöntem olarak organik tarım ortaya çıkmıştır.</p> <p>çeşidi kullanılmaktadır.</p> <p>Organik-inorganik kökenli farklı yetiştirme ortamlarından oluşan topraksız kültür yetiştiriciliğinin ise gerek topraktan kaynak lanan hastalık, nemlendirme ve yabancı ot çıkışının olmaması gerekse erkenci olması açısından dolayı özellikle topraktan buluşık old uğu alanlarda organik topraklı tarım alternatif bir yöntem olarak değerlendirilmeye başlanmıştır.</p> <p>Sayıları azalarak günümüze kadar gelen ve tahminen 2500 adet bir ağaç varlığı ile 200 ton üretimin kurulması, son yıllarda bu ç eşit üzerinde yapılan çalışmalara bağlı olarak klon anaçlar üzerine esli fidanlarla kapama bahçelerinin yapıldığı görülmektedir.</p> <p>Tarım yapılabilecek toprakların azalması, kullanılabılır su kaynaklarının çeşitli nedenlerle yararlanılamazlık duruma gelmesi, a rıtın nüfusla birlikte değerlendirildiğinde, sürdürülebilir nitelikte, sağlıklı, güvenli, yeterli gıda üretimi ve kalitesi bir y an sıra ortam oluşturmada büyük yararları bulunan yenice ilkesi yöre halkı tarafından da tüketim amacı kullanılarak organik ürün yetiştirilmesi ve organik bitkisel ürünlerin geliştirilmesi için araştırılmasıdır.</p> <p>Bu çalışmaların devam niteliğinde de İzmir Büyükşehir Belediyesi, tarlalarda üretilen ürünlerde hiçbir kimyasal maddenin kullan ılmaması ve kesin önemi içme suyu kaynaklarına sahip olan Tahtalı Havzası’nın kirletilmeden korunması amacıyla, havza gövlesi nde organik tarım uygulamaları için çalışmalar yapılmıştır.</p> <p>İklimdeki İşlemsiz Plazma/Organik Emisyon Spektroskopisi (İPE-ÖES) cihazı kullanılarak organik sertifikalı kuruyemişlerin kurşun (Pb), kadmium (Cd), bakır (Cu), nikel (Ni), çinko (Zn), krom (Cr), civa (Hg), demir (Fe), kalay (Sn) ve arsenik (As) element d üzeyleri ölçülmüştür.</p> <p>Bu çalışmada kapsama alanı en fazla olan 9 farklı arazi ürtüğü sınıfa ait alanlarda topraklardan toplamda 237 yizye toprak örn eği alınarak örneklerin suya çözünebilir baskın miktarı bulunan anyon ve katyon içerikleri için kromatografisi ile belirlenmiş tir.</p> <p>Gelecekte, Doğu Anadolu Bölgesi koşullarında; mevcut nüfusun istihdamını korumak, köyden şehre olan kontrolüz dışarı insan göçü nü önlemek, geleceğin, besin ambarı olabilecek dinlenmiş ve uzun yıllar kullanılabılır doğal olarak kendini yenilemiş meşarları n, organik süt koyunu yetiştiriciliği ile yemden üretime hızla dönüşümü sağlanabilir.</p> <p>Diğer yandan köyde bulunan büyük hayvanlarda bulunan venise ilkesi yöre halkı tarafından da tüketim amacı kullanılarak organik ürün yetiştirilmesi ve organik bitkisel ürünlerin geliştirilmesi için araştırılmasıdır.</p> <p>Çay, fındık hatta kivi meyvesinin yanında bölge tarımına katkı sağlayabilecek yeni meyve türleri veya bölgede var olan ancak de ğeri bilinmeyen yöresel meyve türlerinin bölgedeki performanslarının ortaya konulması veya mevcut meyvelerin modern teknikler ile ltimde organik olarak yetiştiriciliğinin çiftçilere gösterilmesi amacıyla Hahşuğu Vakfı tarafından 2004 yılında başlatılan bir çalışma ile Trabzon İli Hayrat İlçesinde 11 farklı meyve türü (maviyemis, kivi, ahududu, böğürtlen, frenk üzümü, Bektaş üzümü, kokulu kara üzüm, karayemiş, turnayemiş, Trabzon hurması ve cilek) ve bu türleri ait 49 farklı çeşit ve/veya tip ile 25 dekar meyve bahçesi tesis edilmiştir.</p> <p>Sonuç olarak; İsa mahallinin literatür analizine göre aşağıda ifade edilen maddeler çerçevesinde organik tarımın kısırlı kalkınma ilişkisinin ele alınabileceği noktalar ortaya konulmuştur: (1) organik tarımın ekonomik çarpan etkisi ve kısırlı istidatın yaratma potansiyeli vardır; (2) organik ürün sertifikasına sahip olan üreticiler bulunmaktadır; (3) organik ürün sertifikasına sa hip olmayan diğer üreticilerin kayıtlarında daha fazla ürünün üretilmesi için etmektedir; (4) Organik tarım konvansiyonel tarım göre de ha sürdürülebilir bir tarımsal üretim şeklidir; (5) Organik tarım, çevre dostu bir uygulamadır; (6) Hükümet Dışı Organizasyonla r (HDO) organik ürün yetiştirilmesinde pazarlamasına kadar pek çok konuda organik üretime olumlu katkı sunmaktadır.</p> <p>Gübre uygulaması: A ve C’de interaktyonunda her iki yıldı da bakır (sırasıyla, 9.3 ve 11.0 mg), baskın katkı tane sayısı (sıra sıyla, 11.7 ve 38.0 adet), 1000 tane agrilajı (sırasıyla, 37.5 ve 38.8 g), tane verimi (sırasıyla, 373.5 ve 384.7 kg/da) ve pro tein oranı (sırasıyla, % 12.6 ve 12.5) bakımından en yüksek değerler geleneksel gübrelere uygulamasında Altay-2000 cesidinden, hektolitre ağırlığı bakımından en yüksek değer birinci yıl 77.0 kg ile geleneksel gübrelere uygulamasında Altay-2000 cesidinde n, ikinci yıl 78.7 kg ile yem geleneksel gübrelere uygulamasında Sulten cesidinden, 82’de baskın katkı (sırasıyla, 37.5 ve 38.6 adet) bakımından ise en yüksek değer deniz yosunu uygulamasında Vildaz cesidinden elde edilmiştir.</p> <p>Araştırma, bu bağlamda bölgenin organik tarım ve eko-turizm için uygunluğu ile çevreye daha az zarar veren ancak 37.5 ile 38.6 arasında artıran bir uygulamaya 1316 tünasıya dönüştürülmüştür.</p> <p>Toplan çok yönlü değerlendirilme sonucunda, "Organik Ürünler" arasında bitkisel atıklardan elde edilen kompostun ikinci d ouz (40ton.ha-1) yeşil gübre bitkisi kombinasyonu (K2+v) uygulaması, ürün verim ve kalitesi açısından her iki yıldı da bir ç uaplarda yer aldığındandır; toprakta nitrat birikimi açısından daha alt gruplarda tespit edildiğinden ve ekonomik yönden de düşük m aliyeti nedeniyle en plano seçilmiştir.</p> <p>Diğer taraftan son dönemlerde yaşanan gelişmeler, organik tarımın yoksulluk, açlık, su kaynakları, sürdürülebilir üretim ve tük etim, iklim değişikliği, ekosistem ve biyolojik çeşitlilik ile ilgili hedeflere ulaşmak için önemli bir role sahip olduğunu gös tirmektedir.</p> <p>Son dönemlerde dünyada özellikle gelişmiş ülkelerde sağlıklı ürün olarak ciddi şekilde rağbet gören ve ülkemizde de bilincel ve satın alma gücü fazla olan tüketicilerin tarafından tercih edilen organik hayvansal ürünler dikkate alındığında özellikle Doğu An adolu Bölgesi koyunculüğünün mevcut yapısı ve üretim biçiminin koyunculüğün bölgedeki en önemli sorunları olan hijyen ve sağlık sorunlarının çözümüne il biriktire çok önemli bir organik koyunculuk potansiyeline sahip olduğunu gözlemlemek mümkündür.</p> <p>Sağlıklı ilgili faydaların bu denli önem arz ettiği ve tüketicilerin özellikle tarımsal kimyasalların insan sağlığı üzerindeki zararlı etkilerinden kaçınma isteklerinin ciddi boyutlara ulaştığı bir ortamda, organik gıda satın almak isteyen ve önemli faktörün, doğal kaynakları ileriiki nesiller için koruma istediğini olması, çalışmaya dahil olan örneklem grubunun üçte iki (altıru stic) yönünün çok güçlü olduğunun bir kanıtı olarak görülmektedir.</p> <p>tesvik edilmelidir.</p> <p>Ayrıca yüksek kesimdeki tarım alanlarının, daha az kimyasal girdi kullanımını nedeniyle organik tarıma daha uygun olduğu şüphelen ilir.</p> <p>Gerek üreticiler gerekse tüketiciler organik ürünler konusunda yeterli bilgiye sahip değildir.</p> <p>Bu konuda yeni çalışmalara ihtiyaç vardır.</p> <p>Yürütülen bu araştırmada de benzer bulgular elde edilmiştir.</p> <p>Türkiye organik hayvancılık açısından önemli bir potansiyele sahiptir.</p> <p>Araştırma devam etmektedir.</p> <p>Yani bir ürün organik olabilir.</p> <p>Bu konuda en güçlü alternatif ise organik tarım olarak düşünülmektedir.</p> <p>Yani bir ürün organik olabilir.</p>
	TextRank	
	LexRank	
Abstract		
	LSA	
	Luhn	
	TextRank	
	LexRank	
Conclusion		
	LSA	
	Luhn	

Table 4. Cosine similarity ratios for data sets

Dataset	Variable	Algorithms		Cosine Similarity (%)
Agriculture	Abstract	TextRank	LexRank	94.18%
		TextRank	LSA	97.81%
		TextRank	Luhn	94.68%
		LexRank	LSA	93.20%
		LexRank	Luhn	92.85%
	Conclusion	LSA	Luhn	93.17%
		TextRank	LexRank	96.20%
		TextRank	LSA	97.50%
		TextRank	Luhn	97.01%
		LexRank	LSA	95.01%
Organic Farming	Abstract	LexRank	Luhn	93.97%
		LSA	Luhn	97.99%
		TextRank	LexRank	93.38%
		TextRank	LSA	97.71%
		TextRank	Luhn	93.45%
	Conclusion	LexRank	LSA	94.61%
		LexRank	Luhn	92.79%
		LSA	Luhn	95.76%
		TextRank	LexRank	93.96%
		TextRank	LSA	98.19%
	Conclusion	TextRank	Luhn	96.57%
		LexRank	LSA	94.01%
		LexRank	Luhn	91.94%
		LSA	Luhn	97.23%

Table 5. Two-dimensional representation of cosine similarity ratios for datasets



In Table 3, the outputs obtained using automatic text summarization methods for the summary variable of the organic agriculture dataset range from [285-301] characters. The number of words for the summary generated by the TextRank algorithm was set to 300. The resulting summary output consisted of 6 sentences and 300 characters. The number of sentences for the summary generated by the LexRank algorithm was set to 27. The resulting summary output consisted of 21 sentences and 301 characters. The number of sentences for the summary generated by the LSA algorithm was set to 8. The resulting summary output consisted of 8 sentences and 291 characters. The number of sentences for the summary generated by the Luhn algorithm was set to 3. The resulting summary output consisted of 3 sentences and 285 characters.

When the cosine similarities of the outputs for the summary variable of the organic agriculture dataset shown in Table 4 were examined, it was observed that the LSA algorithm showed a high degree of similarity with the other algorithms. The similarity rates reached 97.71% between the TextRank and LSA algorithms and 94.61% between the LexRank and LSA algorithms.

According to Table 3, the outputs obtained using automatic text summarization methods for the conclusion variable of the organic agriculture dataset range from [159-235] characters. The number of words for the summary to be generated using the TextRank algorithm was set to 250. The resulting summary output consisted of 5 sentences and 235 characters. The number of sentences for the summary to be generated using the LexRank algorithm was set to 20. The resulting summary output consisted of 17 sentences and 203 characters. The number of sentences for the summary to be generated using the LSA algorithm was set to 5. The resulting summary output consisted of 5 sentences and 215 characters. The number of sentences for the summary to be generated using the Luhn algorithm was set to 1. The resulting summary output consisted of 1 sentence and 159 characters.

When the cosine similarities of the outputs for the conclusion variable of the organic agriculture dataset shown in Table 4 were examined, high similarity rates were observed between the TextRank and LSA algorithms 98.19%, between the LSA and Luhn algorithms 97.23% and between the TextRank and Luhn algorithms 96.57%.

Additionally, in Table 5, the cosine similarity measurements of all algorithms were visualized on a two-dimensional plane, and the algorithms showing high similarity with each other were marked on the two-dimensional plane using a Venn diagram.

4. Discussion

Text summarization is a text mining application that enables the presentation of essential information from small, medium, and especially large-volume text datasets in the form of short, highly representative and

meaningful summaries. Rapid technological advancements in this field have allowed for the development of various statistical, graphical and deep learning-based methods. This section will primarily examine studies involving extractive text summarization techniques and the findings obtained, along with the results of our study, will be evaluated.

Alsawi and Taşçı (2024) developed a new extractive summarization model combining the PageRank algorithm with word embedding. The proposed model generates summaries by calculating sentence importance scores using both graphic-based and semantic-based features. The results showed that it provided higher performance in Arabic texts compared to traditional extractive approaches. Gonzalez et al. (2023) proposed an extractive summarization model based on the attention mechanism. The model demonstrated higher summary quality compared to classical graphic-based methods by learning inter-sentence relationships and producing more coherent and contextual summaries. Srividya et al. (2022) developed a hybrid model using extractive summarization techniques such as Luhn and TextRank, as well as abstract summarization techniques such as the Pegasus model. This model achieved higher success rates compared to other models, particularly in terms of Rouge scores. Abu Nada et al. (2020) introduced a general-purpose architecture based on NLG and NLU that they proposed for summarizing Arabic documents. This study demonstrates the effectiveness of using the AraBERT model and clustering techniques together for extractive text summarization. The maximum independent cluster-based method proposed by Uçkan and Karıcı (2020) offers a new approach in the field of extractive text summarization. This method has achieved successful results in evaluations using Rouge evaluation metrics and has been effective in eliminating various text irregularities. Mahajani et al. (2019) compared existing summarization systems and highlighted the advantages of extractive and abstraction techniques. They stated that both techniques are effective in different use cases and have certain limitations.

In this study TextRank, LexRank, LSA and Luhn algorithms summarized texts using different approaches. The summary outputs produced by the algorithms were compared in two different datasets. While the summaries produced by TextRank, LSA and Luhn algorithms were found to be similar, the summaries produced by the LexRank algorithm were found to be different from these three. In particular, since the similarity rates of TextRank and LSA algorithms were found to be quite high, it can be said that they are summarization methods that can be used interchangeably.

The research results are consistent with recent studies reporting that extractive text summarization methods have strong representational capabilities in short text datasets. In particular, Bagheri Nezhad et al. (2025) showed that classical extractive methods are competitive in terms of summary quality, but newer approaches offer

advantages in terms of fairness. Evgin et al. (2025) reported that combining extractively selected important sentences with abstractive models increased summary performance. These results are also consistent with recent studies reporting that graphic-based and statistical methods have similar representational power in short texts (Luo et al., 2024). However, the literature emphasizes that hybrid approaches and deep learning models are more successful with large and complex datasets (Srividya et al., 2022; Gonzalez et al., 2023; Alselwi and Taşçı, 2024), but classical extractive methods remain competitive with short and limited datasets (Mahajani et al., 2019; Zhang et al., 2025). This situation shows that dataset size and text length are among the key factors determining the performance of summarization methods.

Further studies could evaluate the performance of extractive and abstractive text summarization methods on small, medium and large-scale text datasets. A hybrid approach combining the strengths of both methods could be proposed. Furthermore, similar studies could be conducted in different fields and/or languages to measure the performance efficiency of the algorithms. With these suggestions, we can better utilize the power of algorithms in text summarization processes. These studies help us understand the diversity of text summarization techniques and how these techniques perform on different datasets.

Author Contributions

The percentages of the authors' contributions are presented below. All authors reviewed and approved the final version of the manuscript.

	E.T.	M.M.
C	50	50
D	50	50
S	0	100
DCP	70	30
DAI	70	30
L	70	30
W	80	20
CR	20	80
SR	70	30
PM	40	60
FA	0	0

C= concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The authors declared that there is no conflict of interest.

Ethical Consideration

Ethics committee approval was not required for this study because there was no study on animals or humans.

Acknowledgements

This research is part of the doctoral dissertation of Ebru Temizhan, conducted under the supervision of Prof. Dr. Mehmet Mendes. Ebru Temizhan's doctoral thesis was supported between 2018 and 2022 under the YÖK 100/2000 Doctoral Scholarship Program in the field of Food Technologies: Organic Agriculture. Between 2022 and 2024, the research was further supported by the TÜBİTAK 2211-C Domestic Doctoral Scholarship Program for Priority Areas in Big Data and Data Analytics. We would like to thank the Council of Higher Education (YÖK) and the Scientific and Technological Research Council of Türkiye (TÜBİTAK) for their support in providing doctoral scholarships.

References

- Abu Nada, A. M., Alajrami, E., Al-Saqqa, A. A., & Abu-Naser, S. S. (2020). Arabic text summarization using arabert model using extractive text summarization approach. *International Journal of Academic Information Systems Research*, 4(8), 6–9.
- Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J., & Kochut, K. (2017). A brief survey of text mining: Classification, clustering and extraction techniques. *arXiv*. <https://doi.org/10.48550/arXiv.1707.02919>
- Alselwi, G., & Taşçı, T. (2024). Extractive Arabic text summarization using PageRank and word embedding. *Arabian Journal for Science and Engineering*, 49, 13115–13130. <https://doi.org/10.1007/s13369-024-08890-1>
- Bagheri Nezhad, S., Bandyapadhyay, S., & Agrawal, A. (2025). Fair summarization: Bridging quality and diversity in extractive summaries. In *Proceedings of the 2025 Conference on Computational Natural Language Processing* (pp. 22–34). Association for Computational Linguistics.
- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 394–398.
- Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22, 457–479.
- Evgin, E., Karadeniz, İ., & Yıldız, O. T. (2025). MetninOzU at BioLaySumm2025: Text summarization with reverse data augmentation and injecting salient sentences. In *Proceedings of the 24th Workshop on Biomedical Language Processing* (pp. 179–184). Association for Computational Linguistics.
- Gonzalez, J. A., Segarra, E., Garcia-Granada, F., Sanchis, E., & Hurtado, L.-F. (2023). Attentional extractive summarization. *Applied Sciences*, 13(3), 1458. <https://doi.org/10.3390/app13031458>
- Jupyter Notebook. (2024). *Jupyter Notebook* (Version 6.0.3) [Software]. <https://www.anaconda.com/download> (Erişim tarihi: 13 Ocak 2024).
- Krishna, K., Roy, S., & Iyyer, M. (2021). Hurdles to progress in long-form question answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4940–4957). Association for Computational Linguistics.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An

- introduction to latent semantic analysis. *Discourse Processes*, 25(2-3), 259-284.
- Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of Research and Development*, 2(2), 159-165.
- Luo, M., Xue, B., & Niu, B. (2024). A comprehensive survey for automatic text summarization: Techniques, approaches and perspectives. *Neurocomputing*, 603, Article 128280. <https://doi.org/10.1016/j.neucom.2024.128280>
- Mahajani, A., Pandya, V., Maria, I. P., & Sharma, D. (2019). A comprehensive survey on extractive and abstractive techniques for text summarization. In *Advances in Intelligent Systems and Computing* (pp. 339-351). Springer.
- Mani, I. (2001). *Automatic summarization*. John Benjamins Publishing Company. <https://doi.org/10.1075/nlp.3>
- Mani, I., Klein, G., House, D., Hirschman, L., Firmin, T., & Sundheim, B. (2002). SUMMAC: A text summarization evaluation. *Natural Language Engineering*, 8(1), 43-68. <https://doi.org/10.1017/S1351324901002741>
- Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into texts. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing* (pp. 404-411). Association for Computational Linguistics.
- Qazvinian, V., & Radev, D. R. (2008). Scientific paper summarization using citation summary networks. In *Proceedings of the 22nd International Conference on Computational Linguistics* (pp. 689-696). Coling 2008 Organizing Committee.
- Rao, S. X., Piriyaatamwong, P., Ghoshal, P., Nasirian, S., Mitrović, S., de Salis, E., Wechner, M., Brucker, V., Egger, P., & Zhang, C. (2022). Keyword extraction in scientific documents. *CEUR Workshop Proceedings: Vol. 3361. Proceedings of the 7th Swiss Text Analytics Conference (SwissText 2022)* (ss. 44-55). CEUR-WS.org. <https://doi.org/10.48550/arXiv.2207.01888>
- Rios Gaona, M. A., Gelbukh, A., & Bandyopadhyay, S. (2010). Recognizing textual entailment using a machine learning approach. In *Advances in Soft Computing* (pp. 177-185). Springer.
- Srividya, D. K., Ram, T., Bommuluri, S. K., Basa, V. R., Venkata, V., Kumar, V., Venkata, R., & Chatradi, S. (2022). A hybrid approach for automatic text summarization and translation based on luhn, pegasus, and textrank algorithms. In *2022 International Conference on Smart Generation Computing, Communication and Networking* (pp. 1-8). IEEE.
- Steinberger, J., & Jezek, K. (2004). Using latent semantic analysis in text summarization and summary evaluation. In *Proceedings of the 7th International Conference* (pp. 93-100). ISIM.
- Temizhan, E. (2024). *Application of text mining in agriculture* [Doktora tezi, Çanakkale Onsekiz Mart Üniversitesi]. Fen Bilimleri Enstitüsü.
- Uçkan, T., & Karcı, A. (2020). Extractive multi-document text summarization based on graph independent sets. *Egyptian Informatics Journal*, 21(3), 145-157. <https://doi.org/10.1016/j.eij.2019.12.002>
- Wan, X., & Yang, J. (2008). Multi-document summarization using cluster-based link analysis. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 299-306). Association for Computing Machinery.
- Zhang, H., Yu, P. S., & Zhang, J. (2025). A systematic survey of text summarization: From statistical methods to large language models. *ACM Computing Surveys*, 57(11), 1-41. <https://doi.org/10.1145/3731445>
- Zhang, J., Zhao, Y., Saleh, M., & Liu, P. (2020). PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. In *Proceedings of the 37th International Conference on Machine Learning* (p. 1051). JMLR.