Jurnal Infotel
Vol. 18, No. 1, February 2026, pp. 94–113. doi:10.20895/infotel.v18i1.1416

Research Article

Strategic Planning for Rice Seed Productivity Using Integration of modified TF-IDF and SWOT-QSPM

Enci Mulyani[1], Ridho Ananda[2,*], and Famila Dwi Winati[3]

[1,2,3]Industrial Engineering study Program, Telkom University, Purwokerto Campus, Jl. DI Panjaitan No. 128, Purwokerto 53147, Central Java, Indonesia
[2]Faculty of Information & Technology, Universiti Kebangsaan Malaysia, Lingkungan Ilmu, Bangi 43600, Selangor, Malaysia

*Corresponding email: ridhoa@telkomuniversity.ac.id

Received: July 15, 2025; Revised: January 01, 2026; Accepted: February 17, 2026.

Abstract: The agricultural sector of Indonesia is dependent on the availability of high-quality rice seeds for its functionality. The effective management of these seeds is therefore of paramount importance to ensure the continuity of productivity and the security of food supplies. However, the aspirations of farmers, who are the primary actors, are often ineffective and only available in an unstructured narrative form. This complicates the process of strategic decision-making. The objective of this study is to enhance rice seed productivity by developing a strategy that employs an integrative informatics approach, integrating text mining, SWOT analysis, and the QSPM method. The data was collected via 100 open-ended interviews with farmers and processed through text cleansing, modified TF-IDF weighting, and token classification into SWOT factors. The classification results were then employed to construct IFAS and EFAS matrices, which were used to determine strategic positioning. The utilization of the QSPM matrix facilitated the identification of priority strategies. The analysis indicated that the seed aspect falls into quadrant IV, suggesting a predominance of weaknesses and threats, necessitating a defensive (WT) strategy. The primary strategy identified was the provision of superior seeds that are resistant to extreme weather; this strategy achieved the highest score in the QSPM analysis. The strategy's feasibility level, as validated by three experts, exceeded 83%, thus categorizing it as "highly feasible." The present study concludes that integrating text mining techniques with SWOT-QSPM transforms opinion data into an objective, adaptable, and applicable decision-making strategy based on local data.

Keywords: Rice seeds, SWOT analysis, text mining, TF-IDF, QSPM

---

1 Introduction

Indonesia, as an agricultural nation, is highly dependent on its agricultural sector to sustain its economy and ensure food security. This sector contributes substantially to the Gross Domestic Product (GDP), with rice as the essential product supporting national food requirements [1]. The significant dependence of the population on rice makes the improvement of rice productivity a strategic concern, with seed quality a crucial determinant [2]. However, many farmers continue to use uncertified seeds due to limited availability and higher costs, leading to lower yields [3].

Optimizing rice seed production requires context-specific solutions tailored to farmers' actual needs. The participation of farmers in planning is essential; however, it often proves ineffective due to inadequate institutions, hierarchical communication, and farmers' challenges in systematically expressing their concerns [4, 5]. In Padaherang District, Pangandaran Regency, West Java, strategy design continues to depend on farmers' empirical knowledge, conveyed through unstructured narratives. This local knowledge is abundant in insights. However, these insights are diverse and difficult to assess manually. Consequently, a reliable data analysis technique is essential for deriving these insights, and one such approach is data mining.

Data mining is a multidisciplinary computational methodology that integrates statistical techniques, artificial intelligence, and database management to extract variables [6, 7], identify patterns [8,9], address missing values [10], and function as a recommendation system [11] from extensive datasets. This methodology has been extensively utilized across diverse domains, including consumer behavior analysis [12], financial trend forecasting [13], educational assessment [14–17], metabolic engineering [18, 19], as well as sentiment analysis and document categorization [20]. In agricultural informatics, data mining utilizing machine learning methods is progressively being employed. A Multisensor Machine-Learning Approach (MMLA) has been presented for crop categorization, with the Random Forest algorithm demonstrating superiority in improving prediction accuracy [21]. A Named Entity Recognition (NER) technique based on a Conditional Random Field (CRF) has been developed to build an automated agricultural knowledge base [22]. Additionally, text mining and cluster analysis have been applied to numerous smart farming papers to identify prevailing research trends using TF-IDF and K-means clustering [23].

Text mining is a distinct subset of data mining that concentrates on the analysis of narrative data, utilizing the TF-IDF (Term Frequency-Inverse Document Frequency) method as a principal approach for pinpointing predominant phrases that encapsulate core themes within a text corpus [24]. The application of TF-IDF in this work facilitates the conversion of qualitative interview data from farmers, which was previously challenging to examine manually, into a structured, weighted format. TF-IDF functions by evaluating term frequency (TF) and its scarcity across documents (IDF), indicating that if a term is present in all documents, its TF-IDF value is rendered zero [25]. This arises because TF-IDF is a function intended for classification [26–28]. Meanwhile, TF-IDF is inappropriate for unstructured narratives from farmer interviews. This is because, instead of performing a classification process, the case attempts to gather as much information as possible from them. Nevertheless, TF-IDF's ability to transform qualitative data into quantitative data can indirectly be utilized.

In parallel, strategic planning frameworks such as SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis and the Quantitative Strategic Planning Matrix (QSPM) have been widely adopted to formulate and prioritize development strategies [29]. SWOT facilitates the identification of internal and external factors affecting a system, while QSPM enables quantitative comparison of strategic alternatives [30, 31]. However, traditional SWOT–QSPM applications often rely heavily on expert judgment, which introduces subjectivity and limits transparency [22]. Moreover, the weighting of factors is frequently detached from empirical field data, reducing the representativeness of the resulting strategies [32].

To address these methodological gaps, this study proposes an integrative approach that combines modified TF-IDF–based text mining with the SWOT–QSPM framework for strategic planning in rice seed productivity. The proposed approach transforms unstructured farmer narratives into quantitative weights that directly inform IFAS and EFAS matrices, thereby reducing subjectivity and strengthening empirical grounding. A modified TF-IDF formulation is proposed to capture dominant field issues. This perspective contrasts with many previous studies that rely predominantly on expert assessments or secondary datasets. By grounding analysis in farmers' narratives, the proposed framework supports participatory, evidence-based decision-making aligned with on-the-ground conditions. Therefore, the contributions of this research are delineated as follows:

1. Proposing a modified TF-IDF weighting scheme that preserves the importance of frequently occurring terms in narrative agricultural data.
2. Integrating text mining with SWOT–QSPM to convert qualitative interview data into quantitative strategic priorities.
3. Providing a participatory and data-driven framework for strategic planning in rice seed productivity that reduces subjectivity and enhances transparency.

The remainder of this paper is structured as follows. Section 2 describes the research methodology, including data collection, preprocessing, modified TF-IDF formulation, and integration with SWOT–QSPM. Section 3 presents the results and analysis. Section 4 discusses the findings in relation to previous studies and practical implications. Finally, Section 5 concludes the paper and outlines directions for future research.

2 Research Method

2.1 Population and Sample

The subjects in this study were key agricultural stakeholders in Padaherang Subdistrict, including landowners, farm workers, members of farmer groups and representatives of the agricultural office. The subjects were selected because of their direct participation in rice cultivation activities, which gave them relevant knowledge and experience on issues of rice seed productivity.

The Slovin formula is a widely employed technique in the field of text mining research, which is utilized to determine the sample size of a specific population in an efficient and representative manner [33, 34]. In this study, an error tolerance of 10% was employed, a margin generally considered reasonable for exploration social studies that utilize a narrative approach [35]. This tolerance level was selected to facilitate the exploration of diverse perceptions and experiences of farmers while ensuring the efficient utilization of research resources. The Slovin formula is expressed in Equation 1.

$$n = \frac{N}{1 + N.e^2} \tag{1}$$

In the formula, n refers to the size of the sample. Meanwhile, N denotes the population size. Furthermore, e indicates the tolerance error utilized. This study employed a 10% tolerance error, resulting in a minimum sample size of 100 farmers from a total population of $12,432$ farmers in Padaherang Subdistrict. The selection of a 10% margin of error is justified by the exploratory nature of the study, which aims to capture a broad range of farmers' narratives and perceptions rather than to estimate population parameters. Such an error level is commonly applied in early-stage or social exploration research to balance data diversity and resource constraints. It has been used in similar rice agriculture studies [36].

2.2 The modified TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) is a statistical method in text mining used to measure the importance of a term in a document relative to an entire corpus [37, 38]. This method operates by calculating two main components: TF and IDF. The TF (Term Frequency) value indicates how often a term appears in a single document, normalized by document length to avoid bias. Meanwhile, IDF (Inverse Document Frequency) measures the rarity of a term across the corpus, where terms appearing in many documents are considered less informative [39, 40]. Mathematically, the TF-IDF formula is presented in Equation 2.

$$\begin{aligned} \text{TF-IDF}(t,d) &= \text{TF}(t,d) \times \text{IDF}(t) \\ &= \frac{n_t}{n_d} \times \log\left(\frac{N}{N_t}\right) \end{aligned} \tag{2}$$

Here, TF-IDF(t,d) represents the TF-IDF score for term t in document d. Meanwhile, n_t and n_d denote the frequency of term t and the total number of terms in document d, respectively. Additionally, N and N_t refer to the total number of documents and the number of documents containing term t, respectively.

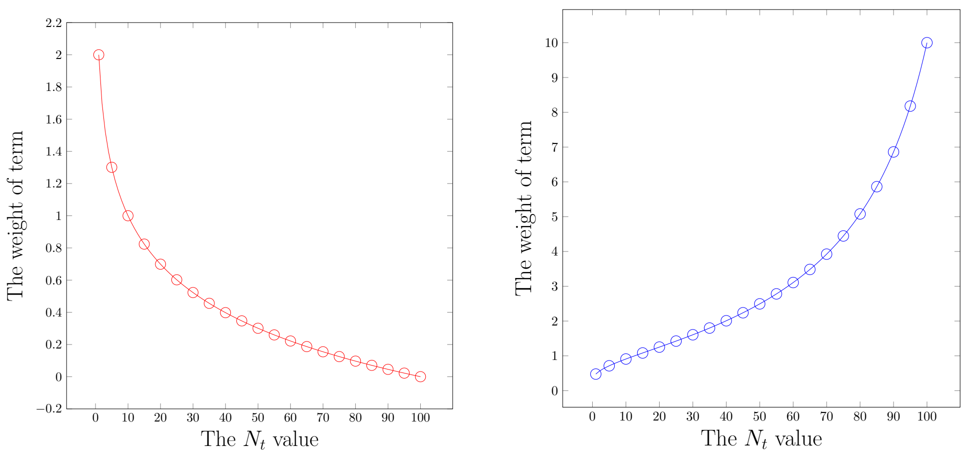
In this study, TF-IDF was modified to address a limitation of the standard TF-IDF. In the standard method, when a term appears in nearly all documents, its IDF is close to zero, indicating it is unimportant. As a result, the essential issues raised by all farmers will be deemed unimportant when assessed using the TF-IDF score, given its formula. This occurs because TF-IDF was designed for document classification tasks [26–28]. Therefore, the modification was necessary because terms consistently mentioned by all respondents indicated critical issues requiring attention. The proposed modified TF-IDF formula is presented in Equation 3.

$$\begin{aligned} \text{Modified TF-IDF}(t,d) &= \text{TF}(t,d) \times \frac{1}{\alpha + \text{IDF}(t)} \\ &= \frac{n_t}{n_d} \times \frac{1}{\alpha + \log\left(\frac{N}{N_t}\right)} \end{aligned} \tag{3}$$

Here, Modified TF-IDF(t,d) represents the modified TF-IDF score for term t in document d, while $\alpha \in \mathbf{R}$ is a weighting parameter. In this study, $\alpha = 0.1$, ensuring the modified

TF-IDF scores fall within the interval [0,1]. The formula was developed based on the prior smoothing approach [41,42]. Through this approach, the modified TF-IDF can identify the words respondents most frequently mention as essential.

Figure 1 compares the weight distributions of terms under standard TF-IDF and the modified TF-IDF across 100 respondents. The figure shows that standard TF-IDF assigns near-zero weights to terms mentioned by all respondents, whereas the modified TF-IDF assigns them maximum weights. Despite the inverse trend, the rate of weight change relative to N_t remains similar between both methods (the rate of term based on N_t around 0.05), indicating that the modification preserves the original weighting dynamics of TF-IDF, although reversing the trend direction.



(a) TF-IDF (b) Modified TF-IDF

Figure 1: Comparison of TF-IDF and Modified TF-IDF weights across 100 respondents.

2.3 The SWOT-QSPM Method

The SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis is a strategic tool used to identify and evaluate internal and external factors of an observed phenomenon [30]. This approach aims to formulate effective strategies by leveraging strengths and opportunities while mitigating weaknesses and anticipating potential threats [43, 44]. However, traditional SWOT analysis is descriptive and qualitative, necessitating supplementary methods for more measurable decision-making. To address this limitation, the Quantitative Strategic Planning Matrix (QSPM) was developed as a quantitative framework that complements SWOT results. QSPM enables researchers or decision-makers to evaluate alternative strategies based on Total Attractiveness Scores (TAS) [31], thereby facilitating objective strategy formulation.

In this study, modified TF-IDF is integrated with SWOT-QSPM to overcome the subjectivity inherent in weighting internal and external factors [22]. Term weights derived from interview data replace conventional respondent-based weighting procedures, ensuring objectivity. The technical integration of modified TF-IDF into SWOT-QSPM follows these steps:

1. Calculating weights for internal/external factors using normalized modified TF-IDF values using Equation 4.

$$w(t) = \frac{\text{TF-IDF}(t, d)}{\sum_{\forall k} \text{TF-IDF}(k, d)}, \qquad \text{where } t, k \in \text{ the same factor} \tag{4}$$

   Where $w(t)$ represents the weight of term t, meanwhile, k denotes terms within the same SWOT factor with t, namely internal (Strengths and Weaknesses) or external (Opportunities and Threats) factors.
2. Prioritizing strategies by pairing SWOT categories. The SO (Strengths-Opportunities) strategy is obtained from the sum of strengths and opportunities weights. Meanwhile, the WO (Weaknesses-Opportunities) strategy is determined by the total value of the Weaknesses and Opportunities weights. Furthermore, the ST (Strengths-Threats) and WT (Weaknesses-Threats) strategies are from the total value of the strengths and threats weights and the weaknesses and threats weights, respectively. Here, the highest-weighted pair determines the priority strategy.
3. Recommending strategies by screening the information of the categories from the highest-weighted pair.
4. Validating strategies using the QSPM method. Priority strategies were identified based on the Sum of Total Attractive Scores (STAS) values for each strategy. The STAS calculation formula in QSPM is presented in Equation 5.

$$\text{STAS}_j = \sum_{\forall t} w(t) \times \text{AS}_{tj} \tag{5}$$

   Where STAS_j represents the STAS value for strategy j. Meanwhile, $w(t)$ denotes the weight of strategic internal and external factors derived from term t. Furthermore, AS_{tj} indicates the Attractive Score obtained from expert assessment of the recommended strategy in relation to internal and external strategic factors within each SWOT category

2.4 Research Flow

The research methodology of this study comprises seven systematic stages. The initial stage involves problem identification and literature review, where research questions related to rice seed productivity are formulated according to SWOT categories (Strengths, Weaknesses, Opportunities, Threats) referring to [45]. The second stage focuses on data collection based on the sample used. Here, the data collection was obtained through interviews, yielding unstructured textual data. Subsequently, the third stage implements standard text mining preprocessing techniques including case folding, tokenizing, filtering, stemming, and handling missing values [46, 47]. The fourth stage calculates term weights using the modified TF-IDF method described in subsection 2.2. Then, the fifth stage develops strategic recommendations through SWOT analysis. The sixth stage employs the quantitative strategic planning matrix (QSPM) method to prioritize the formulated strategies. The final step involves a comprehensive interpretation of the analytical results. The complete research workflow is presented in Figure 2, illustrating the sequential integration of these methodological components.

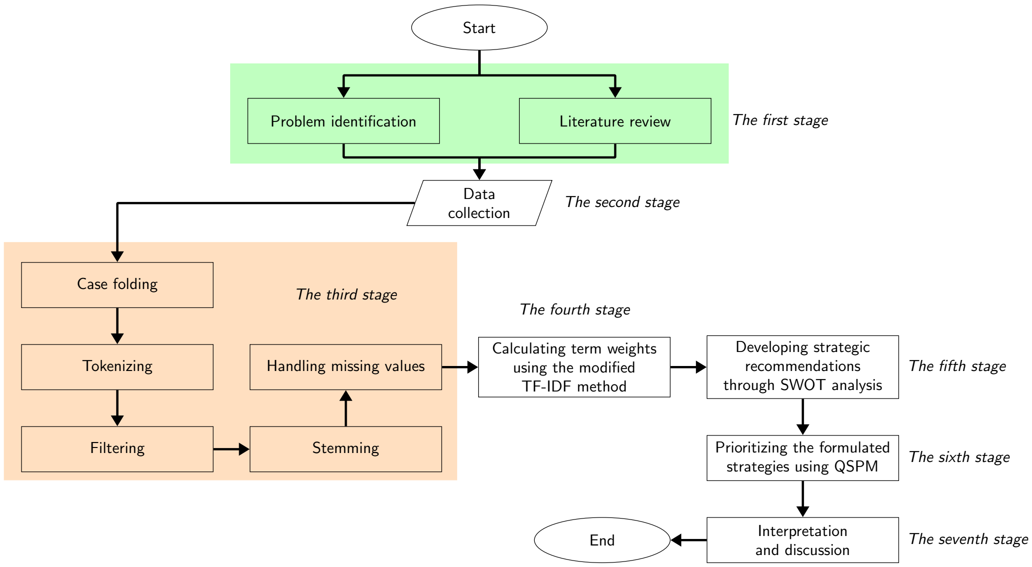


Figure 2: Research Flow of this study

3 Results

3.1 The dataset used

The dataset for this study was collected through interviews with 100 farmers distributed in the Padaherang Subdistrict area, focusing specifically on the productivity of rice seeds. The semi-structured interviews employed pre-designed guiding questions based on the SWOT analysis framework, allowing for flexible exploration of farmers' narrative responses while maintaining alignment with research objectives. As presented in Table 1, the interview questions were systematically organized according to SWOT categories (Strengths, Weaknesses, Opportunities, Threats) following the framework utilized by [45].

Table 1: The study questions

No	Question	Category
1	Menurut Anda, bagaimana kelebihan penggunaan benih yang selama ini digunakan untuk meningkatkan hasil produksi?	Strength
2	Kendala seperti apa yang sering Anda temukan ketika memilih benih padi yang dapat mengurangi/melemahkan kualitas produksi?	Weaknesses
3	Menurut Anda, hal-hal apa saja yang mendukung penggunaan benih padi memberikan hasil yang baik hingga saat ini dan memberikan prospek di masa mendatang?	Opportunity
4	Menurut Anda, masalah seperti apa yang dapat mengancam kualitas benih padi?	Threat

Furthermore, the obtained dataset comprises 400 narrative responses from four SWOT categories. Subsequently, the dataset was analyzed to retrieve insights about rice seed productivity challenges and opportunities. The qualitative nature of these responses required specialized text mining approaches to uncover meaningful patterns and strategic information about the phenomena observed.

3.2 Term weighting using modified TF-IDF

Furthermore, dataset preprocessing is subsequently performed through several methodological steps. The first step, text cleaning and case folding, was performed to standardize the text format and remove irrelevant elements such as numbers, symbols, double spaces, and single characters. Next, tokenizing was applied—splitting the text into tokens using the word_tokenize function from the NLTK library to enable word-level analysis. The resulting tokens were then filtered through stop word removal using the Sastrawi library to eliminate common words that hold little analytical value. The process continued with stemming, which converts derivative words into their root forms to ensure semantic consistency and comparability. Then, the last step was the identification and handling of missing values to maintain analysis quality. Several datasets were found to be empty or resulted in a 'nan' term. This occurred due to respondents' inability to provide narrative information in response to the posed questions. To address this issue, an imputation process was carried out to ensure the dataset remained complete and could be analyzed optimally. The imputation technique applied in this study is mode-based imputation, which involves filling missing values using the most frequently occurring terms within each category. All of these steps were performed sequentially to ensure that the narrative text data became clean, uniform, and ready for further analysis. Table 2 presents several results of the dataset preprocessing.

Table 2: The preprocessing results

No.	Raw Data	Preprocessing results
1	Kualitas benihnya bagus	['kualitas', 'benih', 'bagus']
2	Benih diperoleh dari hasil panen sendiri dan dipilih yang terbaik	['benih', 'hasil', 'panen', 'pilih', 'baik']
3	Benih ditabur merata agar pertumbuhan seragam	['benih', 'tabur', 'rata', 'tumbuh', 'seragam']
4	Menggunakan benih unggul jenis IR 64	['benih', 'unggul', 'jenis', 'ir']
5	Menggunakan benih unggul varietes ciherang	['benih', 'unggul', 'varietes', 'ciherang']
⋮	⋮	⋮
100	Sudah punya kebiasaan dan pengetahuan tentang teknis semai dan tanam benih	['biasa', 'tahu', 'teknis', 'semai', 'tanam', 'benih']

After text preprocessing, term weighting was conducted to assess term significance within the dataset using the modified TF-IDF method. This analysis focused on terms with the highest TF-IDF values that were explicitly related to issues of rice seed productivity, as identified in prior studies. The resulting terms and their weights, categorized by SWOT dimensions, are presented in Figure 3. This analysis yielded 13, 12, 11, and 7 key terms for the Strength, Weakness, Opportunity, and Threat categories, respectively.

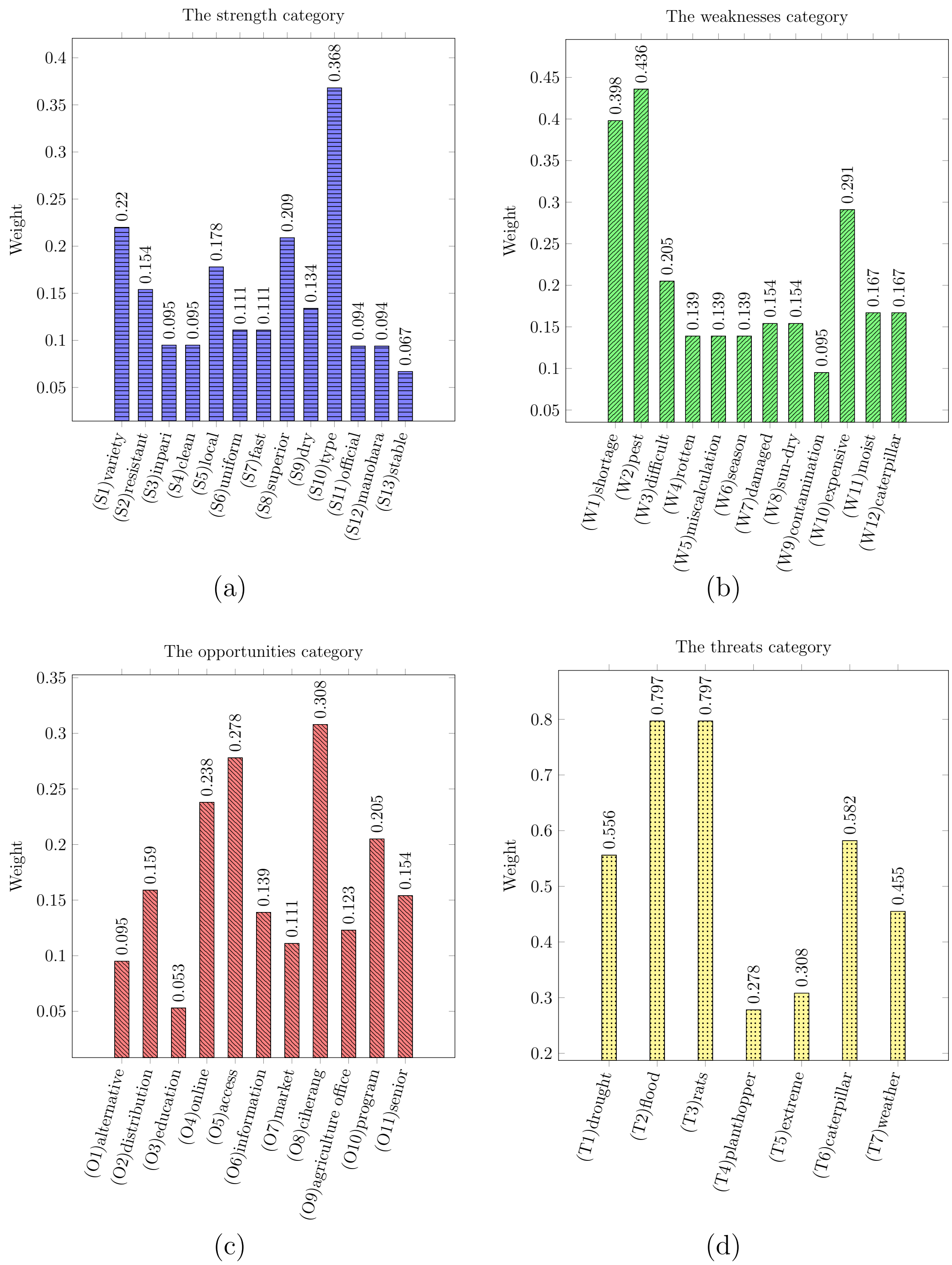


Figure 3: The distribution of the resulting terms and their weights from the modified TF-IDF, categorized into SWOT dimensions, namely (a) Strength, (b) Weakness, (c) Opportunity, and (d) Threat

3.3 Recommended strategies based on SWOT analysis

The analysis process continued with the development of the IFAS and EFAS matrices to get the score of the internal and external factors. The weight of each factor was calculated using the modified TF-IDF values, which reflect the level of importance of each term in the interview narratives. The TF-IDF values were then normalized by dividing them by the total TF-IDF value within each category (IFAS or EFAS), resulting in relative weights ranging from 0 to 1. These weights were used to indicate the relative contribution of each factor in the strategy formulation process. The results of the IFAS-EFAS matrix calculation, along with the relative weights, are presented in Table 3.

Table 3: Weighting of internal and external strategy factors

Internal factors			External factors		
Symbol	TF-IDF	w	Symbol	TF-IDF	w
Strengths			Opportunities		
S1	0.220	0.050	O1	0.095	0.017
S2	0.154	0.035	O2	0.159	0.028
S3	0.095	0.022	O3	0.053	0.009
S4	0.095	0.022	O4	0.238	0.042
S5	0.178	0.040	O5	0.278	0.049
S6	0.111	0.025	O6	0.139	0.025
S7	0.111	0.025	O7	0.111	0.020
S8	0.209	0.047	O8	0.308	0.055
S9	0.134	0.030	O9	0.123	0.022
S10	0.368	0.083	O10	0.205	0.036
S11	0.094	0.021	O11	0.154	0.027
S12	0.094	0.021		Sub Total	0.331
S13	0.067	0.015			
	Sub Total	0.422			
Weaknesses			Threats		
W1	0.398	0.090	T1	0.556	0.099
W2	0.436	0.099	T2	0.797	0.141
W3	0.205	0.046	T3	0.797	0.141
W4	0.139	0.031	T4	0.278	0.049
W5	0.139	0.031	T5	0.308	0.055
W6	0.139	0.031	T6	0.582	0.103
W7	0.154	0.035	T7	0.455	0.081
W8	0.154	0.035		Sub Total	0.0.669
W9	0.095	0.022			
W10	0.291	0.066			
W11	0.167	0.038			
W12	0.167	0.038			
	Sub Total	0.563			

Based on the combination of IFAS and EFAS scores shown in Table 3, the IFAS score difference is –0.14, and the EFAS score difference is –0.34. These differences indicate that the seed management position in the agricultural sector of Padaherang is dominated by negative factors, namely weaknesses and threats. This places the seed aspect in Quadrant IV, which requires strategies focused on minimizing weaknesses and avoiding threats, thus, more defensive strategies or WT strategies, as illustrated in Figure 4. The formulation of WT (Weaknesses–Threats) strategies is presented as follows:

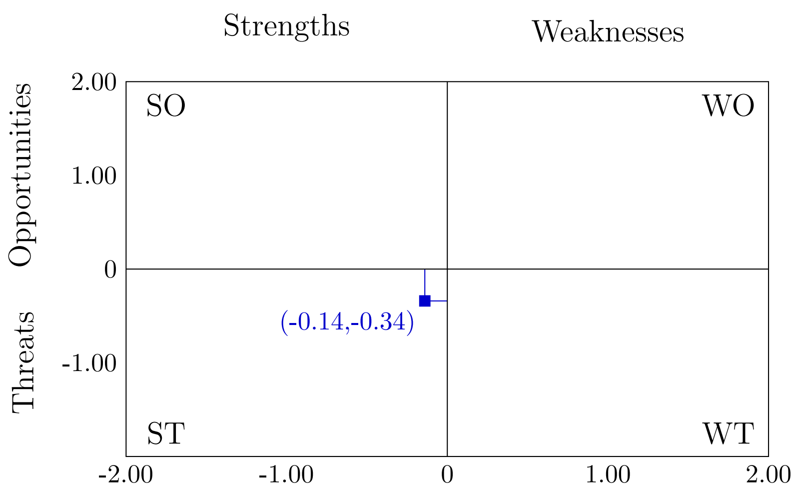


Figure 4: SWOT diagram for seed aspects

1. Conduct training on seed storage to prevent damage and pest attacks and planting seeds at the appropriate time of year (W2, W7 - T1, T2, T3)
2. Introduce safer and longer-lasting seed storage technologies to reduce the risk of damage from humidity and pests (W4, W8, W9, W11, W12 - T5)
3. Provide high-quality seeds resistant to extreme weather to mitigate the impact of drought and pest attacks (W3 - T4, T5, T6, T7)
4. Build a faster and cheaper seed distribution system to ensure timely delivery to farmers and reduce losses due to delays (W1, W5, W10 - T7)
5. Develop seeds that are more resistant to pests such as rats and brown planthoppers through breeding techniques or specific treatments (W6, W3 - T3)
6. Provide safer, quality-controlled storage facilities to ensure seeds remain undamaged during storage (W7, W11 - T5)

These WT strategies focus on addressing internal weaknesses related to seed storage and distribution, while simultaneously responding to external threats such as extreme weather and pest attacks. Approaches such as training, the introduction of storage technology, and timely distribution directly contribute to loss reduction and seed quality improvement. This strategy lays the groundwork for strengthening the seed production system's resilience among farmers.

3.4 Validation of recommended strategies using QSPM

After formulating various strategic alternatives through the SWOT matrix, the next stage is to conduct an analysis using the Quantitative Strategic Planning Matrix (QSPM) to determine the most feasible priority strategy for implementation. QSPM is used to provide a quantitative assessment of each strategy based on the internal and external factors previously identified through the IFAS and EFAS matrices. This process involved evaluations from three agricultural experts from Padaherang District, ensuring an objective assessment by comparing the Attractiveness Scores of each strategy against the weights of the SWOT factors. As a result, strategic decisions are made based on data-driven insights, rather than solely on intuition. Figure 5 demonstrated the sum of total attractive score (STAS) from QSPM for each strategy, where the STAS values of the first, second, third, fourth, fifth, and sixth strategies are 7.16, 7.23, 7.26, 7.18, 6.64, and 7.25, respectively.

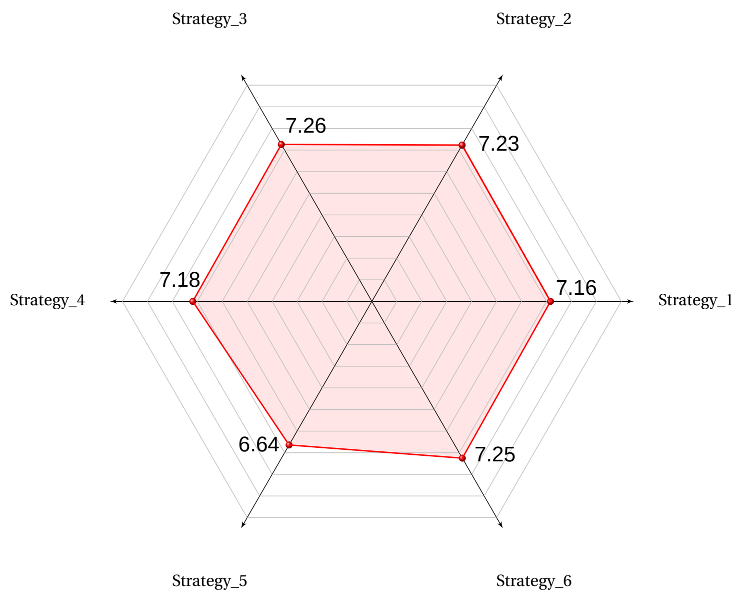


Figure 5: Sum of total attractive score values for each strategy

Based on the QSPM results, the strategy with the highest score is Strategy 3 (score: 7.26), which involves the provision of high-quality seeds resistant to extreme weather and is considered the most important and impactful. This is followed by Strategy 6 with a score of 7.25 and Strategy 2 with a score of 7.23. These three strategies are identified as the top priorities in seed management. Meanwhile, other strategies, such as Strategy 4, Strategy 1, and Strategy 5, are ranked lower, with scores below the top three.

The QSPM-derived strategic priorities are empirically reinforced through direct linkage between each strategy and quantified field conditions. Strategies addressing seed handling and storage (Strategies 1, 2, and 6) are supported by national evidence showing that harvest and postharvest losses in Indonesia reach 20.5% of total rice production, equivalent to economic losses exceeding IDR 15 trillion, primarily driven by suboptimal harvesting, drying, and storage practices [48]. The prioritization of climate-resilient seed provision (Strategy 3) aligns with observed drought stress during the 2023 El Niño event, where the severity of drought in Java averaged -0,84 $\pm$ 0,28 with dry periods lasting up to five consecutive months, compounded by rainfall deficits of 150 to 200 mm per month [49, 50]. Pest-resistance strategies (Strategy 5) are justified by extensive biotic pressure, as pest outbreaks in West Java alone affected more than 5,600 ha of rice fields by stem borers, 4,900 ha by rodents and more than 3,000 ha by bacterial leaf blight within a single planting season [51]. Finally, the emphasis on improving seed distribution efficiency (Strategy 4) is empirically grounded in the dominance of informal seed systems, where 98% of farmers rely on self-saved seeds and only 35% access commercial or formal distribution channels, largely due to delayed or limited availability from official providers [52]. Together, these quantified indicators substantiate the QSPM rankings by demonstrating coherence between expert judgment, farmer narratives, and measurable production constraints.

Furthermore, three experts conducted strategy validation using a Likert scale of 1 to 5 to assess the feasibility of the proposed strategies with respect to field conditions. According to Table 4, the validation results show total scores ranging from 83.77% to 93.02%, which are categorized as "Very Worthy." This assessment reflects a high level of acceptance from the three main stakeholder groups: Landowners, Heads of Farmer Groups, and the District Agriculture Office.

The selection of landowners, farmer groups leaders and representatives of the District Agriculture Office is based on their strategic roles in agricultural decision-making and implementation. Landowners determine production investment and seed adoption decisions, heads of farmer groups coordinate collective farming practices and information dissemination, while agricultural office representatives are responsible for policy execution and technical guidance. Compared to field laborers or farm workers, these stakeholders possess broader decision authority, institutional knowledge, and policy-related insight, making them more suitable for evaluating strategic feasibility at the planning level.

Table 4: Strategy Validation

Stakeholder Identity	Strategy Assessment	Assessment Category
Landowner	0.9302	Very Worthy
Head of Farmer Group	0.8377	Very Worthy
District Agriculture Service	0.8981	Very Worthy

Most of the respondents considered the strategies formulated to be relevant to the field conditions, easy to understand, and realistic to implement. Although the feedback was generally positive, the agriculture office made constructive suggestions, particularly regarding the adaptation of seed varieties to local conditions. In general, this validation reinforces the strength of the strategies as applicable, participatory and grounded in the real needs of farmers.

4 Discussion

4.1 Comparison with Related Studies

This study develops an integrative approach that combines text mining based on TF-IDF, SWOT analysis, and the QSPM method to formulate strategies to systematically improve traditional agricultural productivity. This approach overcomes the limitations of conventional SWOT, which tends to be subjective, by transforming open-ended interview data from 100 farmers into objective quantitative weights through text processing and relative frequency weighting. This conversion supports the construction of IFAS–EFAS matrices and the prioritization of strategies using QSPM, grounded in field data.

Compared to the study by Cheng et al. (2021), which applied change mining to derive SWOT factors from product reviews in a measurable way [29], the approach in this study stands out for applying text mining in the agricultural context and for emphasizing local farmer participation. Meanwhile, Miric et al. (2022) utilized machine learning for automatic classification, and Panoutsopoulos et al. (2022) developed a named entity recognition (NER) model to extract agricultural terms [53, 54]. In contrast, this study emphasizes the transformation of narrative opinions into actionable data that can be used directly in tactical planning. Although Tay (2024) also highlights the need for innovative approaches in SWOT [55], this study excels in efficiency, replicability, and a stronger local context. Thus, it contributes to a new methodological model that can be applied not only in agriculture but also in participatory evidence-based decision-making in other sectors.

4.2 Integration of Text Mining and SWOT-QSPM in Agricultural Context

The method used in this study emphasizes converting qualitative data into quantitative insights through a text-mining approach. The integration of text mining with SWOT-QSPM in this study not only yields more representative strategies but also introduces a new way to quantify qualitative opinion data from farmers. Through preprocessing, modified TF-IDF weighting, and mapping to the SWOT framework, narrative information was successfully converted into a measurable, comparable analytical structure. This makes the strategy formulation process more transparent, data-driven and contextually relevant to the real challenges farmers face on the ground.

4.3 Practical and Policy Implications

The proposed defensive strategies (WT) have several practical implications regarding costs, implementation challenges, and policy alignment. From a cost perspective, adopting superior rice seed varieties that are more resistant to extreme weather conditions requires an initial investment, particularly in seed procurement, storage, and farmer training. However, these costs are relatively lower than the potential losses from crop failure, decreased productivity, and repeated replanting due to climate-related risks.

In terms of implementation, several challenges may arise, including limited access to quality seed storage facilities, varying levels of technical knowledge among farmers, and distribution constraints across farming areas. These challenges highlight the importance of coordination through farmer groups and institutional support from local agricultural authorities to ensure effective dissemination and adoption of the proposed strategies.

From a policy perspective, the recommended strategies are aligned with existing government programs on food security, adaptation to climate change, and the promotion of certified superior seeds. This alignment increases the feasibility of implementation, as the strategies can be integrated into ongoing agricultural extension activities and policy frameworks without requiring major structural changes.

4.4 Replication Potential

The integrative method developed in this study demonstrates strengths in efficiency and flexibility, particularly in handling complex, unstructured, open-ended interview data. With this approach, decision-making is no longer solely dependent on expert judgment but also incorporates the majority's voice through relative-frequency-based weighting and quartile classification. Therefore, this model has the potential to be replicated and further developed as a systematic framework for the design of text-based strategies in other policy sectors and could evolve into a decision-support system based on local data. By making farmers' voices the foundation of strategy, this approach supports applicable and adaptive participatory development practices. The QSPM framework used in this study can be accessed at https://github.com/enci-mulyani003/QSPM-Spreadsheet-Template-for-Text-Mining-SWOT-Integration.

5 Conclusion

This study underscores the strategic role of text mining in the processing of open-ended interview data to support data-driven strategic planning in the agricultural sector. It shows that an integrative informatics-based approach combining text mining, SWOT analysis, and the QSPM method can produce farming strategies that are more systematic, participatory, and grounded in field data. SWOT analysis revealed that the seed management aspect in Padaherang Subdistrict falls within Quadrant IV, indicating a dominance of weaknesses and threats. As a result, WT strategies were formulated as defensive approaches, focusing on improving seed storage systems, distribution mechanisms, and protection against environmental factors and pests.

The top-priority strategy, as identified through the QSPM matrix, is the provision of superior seed varieties resistant to extreme weather conditions, which received the highest score. Expert validation demonstrated very high feasibility, reinforcing that the strategy is relevant, realistic, and practically implementable in the field. The primary strength of this research lies in its use of text mining to transform farmers' narrative input into quantitative data for further analysis within the SWOT-QSPM framework. This process enables the development of a more objective and transparent strategic structure while also incorporating farmer participation as a primary data source. Compared to conventional approaches or similar studies, this model is more efficient in handling unstructured data and can be replicated across other sectors to support text-based decision-making. By anchoring strategic formulation in local perspectives, this research contributes to applied informatics while supporting inclusive and adaptive agricultural development grounded in farmer-driven evidence. Nevertheless, several weaknesses should be acknowledged, including limited sampling coverage, limited stakeholder validation, and the use of a simple mode-based approach to impute missing values, which may affect the natural distribution of responses and the robustness of the findings. These aspects were not fully addressed in the present study due to its exploratory focus and practical constraints related to time, resources, and data availability. Future studies are encouraged to involve larger and more diverse samples, expand stakeholder participation through broader surveys or pilot field testing, and apply more advanced missing-value imputation techniques to improve empirical validity and analytical reliability.

Acknowledgments

The author would like to express sincere gratitude to the farmers in Padaherang Subdistrict for their time and the valuable information they provided. Special thanks are also extended to Telkom University for the academic support and research facilities throughout the study. Appreciation is given to all individuals and parties who contributed to this research, both directly and indirectly.

References

[1] D. Bagus Wiranatakusuma, R. Fairuztama, and A. Aprizal, "Analyzing determinants of economic growth on agricultural sector in Indonesia," *E3S Web of Conferences*, vol. 595, p. 01006, 2024.

[2] T. Wang, Z. Yu, R. Ahmad, S. Riaz, K. U. Khan, S. Siyal, M. A. Chaudhry, and T. Zhang, "Transition of bioeconomy as a key concept for the agriculture and agribusiness development: An extensive review on ASEAN countries," *Frontiers in Sustainable Food Systems*, vol. 6, p. 998594, Nov. 2022.

[3] K. Ilangathilaka, L. Rupasena, and S. Fernando, "Factors Affecting Rice Farmers' Choice of Formal Seed," *Applied Economics & Business*, vol. 5, pp. 51–60, June 2021.

[4] A. De Sherbinin, A. Bowser, T.-R. Chuang, C. Cooper, F. Danielsen, R. Edmunds, P. Elias, E. Faustman, C. Hultquist, R. Mondardini, I. Popescu, A. Shonowo, and K. Sivakumar, "The Critical Importance of Citizen Science Data," *Frontiers in Climate*, vol. 3, p. 650760, Mar. 2021.

[5] P. Dipela, "The Realities and Barriers of Community Participation in Rural Community Development Projects: none," *Community Development Journal*, vol. 8, pp. 379–399, Dec. 2024.

[6] R. Ananda, D. Rachmawaty, B. Pratikno, O. A. Hamid, and M. B. Mohd Amin, *Improved Clustering-Based Feature Selection Using Feature Extraction Based on Principal Component Analysis*, vol. 2329, p. 15–38. Cham: Springer Nature Switzerland, 2025.

[7] R. Ananda, A. R. Dewi, M. B. Mohd Amin, M. Huda, and G. Gushelmi, "Unsupervised feature selection based on self-configuration approaches using multidimensional scaling," *Jambura Journal of Mathematics*, vol. 5, p. 351–362, Aug. 2023.

[8] R. Ananda and A. Prasetiadi, "Hierarchical and k-means clustering in the line drawing data shape using procrustes analysis," *JOIV: International Journal on Informatics Visualization*, vol. 5, p. 306, Sept. 2021.

[9] R. Ananda and A. Prasetiadi, "Classification based on configuration objects by using procrustes analysis," *JURNAL INFOTEL*, vol. 13, p. 76–83, May 2021.

[10] R. Ananda, A. R. Dewi, and N. Nurlaili, "A comparison of clustering by imputation and special clustering algorithms on the real incomplete data," *Jurnal Ilmu Komputer dan Informasi*, vol. 13, p. 65–75, July 2020.

[11] R. Ananda, M. Z. Naf'an, A. B. Arifa, and A. Burhanuddin, "Recommendation system for specialization selection using k-means density canopy," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, p. 172–179, Feb. 2020.

[12] S. Sinaga, R. Ananda, H. Q. Karima, and A. M. Tazuddin, "Marketing analysis of shoe products using principal coordinates analysis and k-means clustering based on the marketing mix at bintang sepatu purwokerto msme," *Jurnal Teknik Informatika (Jutif)*, vol. 6, p. 1405–1418, June 2025.

[13] N. Ozbaltan, "Applying machine learning to audit data: Enhancing fraud detection, risk assessment and audit efficiency," *EDPACS*, vol. 69, no. 9, p. 70–86, 2024.

[14] R. Ananda, "Silhouette density canopy k-means for mapping the quality of education based on the results of the 2019 national exam in banyumas regency," *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 5, p. 158–168, Dec. 2019.

[15] R. Liu, "Data analysis of educational evaluation using k-means clustering method," *Computational Intelligence and Neuroscience*, vol. 2022, p. 1–10, 2022.

[16] I. K. Khan, H. B. Daud, N. B. Zainuddin, R. Sokkalingam, M. Farooq, M. E. Baig, G. Ayub, and M. Zafar, "Determining the optimal number of clusters by enhanced gap statistic in k-mean algorithm," *Egyptian Informatics Journal*, vol. 27, p. 100504, 2024.

[17] R. Ananda and A. Z. Yamani, "Determination of initial k-means centroid in the process of clustering data evaluation of teaching lecturers," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 4, p. 544–550, June 2020.

[18] K. M. Daud, R. Ananda, S. Zainudin, and C. W. Howe, "Optimizing the production of valuable metabolites using a hybrid of constraint-based model and machine learning algorithms: A review," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 10, 2023.

[19] R. Ananda, K. M. Daud, and S. Zainudin, "A review of advances in integrating gene regulatory networks and metabolic networks for designing strain optimization," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, p. 102120, July 2024.

[20] K. Alemerien, A. Al-Ghareeb, and M. Z. Alksasbeh, "Sentiment analysis of online reviews: A machine learning based approach with tf-idf vectorization," *Journal of Mobile Multimedia*, vol. 20, p. 1089–1116, Sept. 2024.

[21] A. Reyana, S. Kautish, P. M. S. Karthik, I. A. Al-Baltah, M. B. Jasser, and A. W. Mohamed, "Accelerating Crop Yield: Multisensor Data Fusion and Machine Learning for Agriculture Text Classification," *IEEE Access*, vol. 11, pp. 20795–20805, 2023.

[22] Z. Wang, Z. Cheng, C. Guan, and H. Han, "Intelligent Information extraction algorithm of Agricultural text based on Machine Learning method," *Journal of Physics: Conference Series*, vol. 1952, p. 022073, June 2021.

[23] S. Sharma, C. Sharma, E. Asenso, and K. Sharma, "Research Constituents and Trends in Smart Farming: An Analytical Retrospection from the Lens of Text Mining," *Journal of Sensors*, vol. 2023, p. 6916213, Jan. 2023.

[24] C. Li, W. Li, Z. Tang, S. Li, and H. Xiang, "An Improved Term Weighting Method Based on Relevance Frequency for Text Classification," Aug. 2021.

[25] M. Liang and T. Niu, "Research on Text Classification Techniques Based on Improved TF-IDF Algorithm and LSTM Inputs," *Procedia Computer Science*, vol. 208, pp. 460–470, 2022.

[26] L. Xiang, "Application of an Improved TF-IDF Method in Literary Text Classification," *Advances in Multimedia*, vol. 2022, pp. 1–10, May 2022.

[27] Z. Jiang, B. Gao, Y. He, Y. Han, P. Doyle, and Q. Zhu, "Text Classification Using Novel Term Weighting Scheme-Based Improved TF-IDF for Internet Media Reports," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–30, Mar. 2021.

[28] S. Jain, S. K. Jain, and S. Vasal, "An Effective TF-IDF Model to Improve the Text Classification Performance," in *2024 IEEE 13th International Conference on Communication Systems and Network Technologies (CSNT)*, (Jabalpur, India), pp. 1–4, IEEE, Apr. 2024.

[29] L.-C. Cheng, K. Chen, M.-C. Lee, and K.-M. Li, "User-Defined SWOT analysis – A change mining perspective on user-generated content," *Information Processing & Management*, vol. 58, p. 102613, Sept. 2021.

[30] M. I. Lufti, R. Mu'min, and J. Haekal, "Determination of digital marketing strategy with application of quantitative strategic planning matrix (QSPM) in enterprise resource planning (ERP) system company's snell X," *Jurnal PASTI (Penelitian dan Aplikasi Sistem dan Teknik Industri)*, vol. 16, p. 27, Apr. 2022.

[31] Yodfiatfinda and D. Safitri, "Development Strategy Of Agro-tourism For Sustainability Of Agricultural Business Entity: Case Study Of Sawangan Farm In Depok, West Java," *IOP Conference Series: Earth and Environmental Science*, vol. 1364, p. 012006, June 2024.

[32] V. Jain, P. Ajmera, and J. P. Davim, "SWOT analysis of Industry 4.0 variables using AHP methodology and structural equation modelling," *Benchmarking: An International Journal*, vol. 29, pp. 2147–2176, Aug. 2022.

[33] A. Padalia and T. Natsir, "End-user computing satisfaction (eucs) model: Implementation of learning management system (lms) on students satisfaction at universities," *International Journal of Environment, Engineering and Education*, vol. 4, p. 100–107, Dec. 2022.

[34] A. Triayudi, W. O. Widyarto, and V. Rosalina, "CLG Clustering for Mapping Pattern Analysis of Student Academic Achievement," 2020.

[35] E. Pradana and M. Sitorus, "The effect of village leadership on community participation in village development in village keramat gajah," *Journal of Economics and Business (JECOMBI)*, vol. 2, p. 97–101, Sep. 2020.

[36] I. Sokra, S. Somaly, H. Meta, S. Horn, and C. Molikoy, "Factors affecting rice production: A systematic review," p. 863670 Bytes, 2025.

[37] K. SPARCK JONES, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 28, p. 11–21, Jan. 1972.

[38] K. Spärck Jones, "Idf term weighting and ir research lessons," *Journal of Documentation*, vol. 60, p. 521–523, Oct. 2004.

[39] A. Ashutosh, A. Sharma, and M. A. Beg, "Strategic analysis using SWOT-AHP: a fibre cement sheet company application," *Journal of Management Development*, vol. 39, pp. 543–557, Feb. 2020.

[40] A. Rakhman, T. J. R. Sitinjak, and T. Sitinjak, "Optimization of Warehouse Selection with SWOT and AHP Methods in the Pulogadung Industrial Area," *International Journal of Social Science and Business*, vol. 7, pp. 817–823, Nov. 2023.

[41] A. Alshehri and A. Algarni, "TF-TDA: A Novel Supervised Term Weighting Scheme for Sentiment Analysis," *Electronics*, vol. 12, p. 1632, Mar. 2023.

[42] R. Diehl Martinez, Z. Goriely, A. Caines, P. Buttery, and L. Beinborn, "Mitigating Frequency Bias and Anisotropy in Language Model Pre-Training with Syntactic Smoothing," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, (Miami, Florida, USA), pp. 5999–6011, Association for Computational Linguistics, 2024.

[43] Kresnawati, R. Nurazi, and R. Indriani, "SWOT Analysis and QSPM as a Strategy for Developing Village Owned Business Units (BUMDes) Tanjung Beringin Village," *Journal of Economics, Finance And Management Studies*, vol. 07, May 2024.

[44] R. N. Safitri, S. Arbainah, and T. D. Karyanti, "Formulation of business strategy using swot method and quantitative strategic planning matrix (QSPM): (case study Surya Boutique Hotel Semarang)," *Applied Accounting and Management Review (AAMAR)*, vol. 2, p. 59–69, May 2023.

[45] A. Singh, "Judicious and optimal use of water and land resources for long-term agricultural sustainability," *Resources, Conservation & Recycling Advances*, vol. 13, p. 200067, May 2022.

[46] W. Bourequat and H. Mourad, "Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine," *International Journal of Advances in Data and Information Systems*, vol. 2, pp. 36–44, Apr. 2021.

[47] D. Cenitta, R. V. Arjunan, and P. K. V, "Ischemic Heart Disease Multiple Imputation Technique using Machine Learning Algorithm," *Engineered Science*, 2022.

[48] A. F. Falatehan, Y. Syaukat, H. Hastuti, and N. Nasrullah, "Paddy Loss and Its Implication to Fertilizer Subsidy in Indonesia," *HAYATI Journal of Biosciences*, vol. 28, p. 73, Jan. 2021.

[49] L. Q. Avia, E. Yulihastin, M. H. Izzaturrahim, R. Muharsyah, H. Satyawardhana, I. Sofiati, E. Nurfindarti, and Gammamerdianti, "The spatial distribution of a comprehensive drought risk index in java, indonesia," *Kuwait Journal of Science*, vol. 50, p. 753–760, Oct. 2023.

[50] M. W. Hassan, M. A. Hashmi, G. Sarwar, Z. Mehmood, W. Saleem, and M. A. Farooqi, "Damage assessment of stored grain pests against rice grains types and wheat," *International Journal of Tropical Insect Science*, vol. 43, p. 35–41, Oct. 2022.

[51] L. Aprilia, Y. Koesmaryono, and S. Priyambodo, "Potential attack of rice field rat (rattus argentiventer) on rice crops based on climate factors in karawang regency, west java, indonesia," *JPT Jurnal Proteksi Tanaman (Journal of Plant Protection)*, vol. 8, p. 63–77, Dec. 2024.

[52] Y. Surdianto, N. Sutrisna, A. Nurawan, K. K. Hamdani, W. Wiratno, E. Erythrina, and D. Sugandi, "Study of rice (oryza sativa) seed breeding based on farmer groups in pangandaran regency, west java province, indonesia," *IOP Conference Series: Earth and Environmental Science*, vol. 1364, no. 1, p. 012020, 2024.

[53] M. Miric, N. Jia, and K. G. Huang, "Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents," *Strategic Management Journal*, vol. 44, pp. 491–519, Feb. 2023.

[54] H. Panoutsopoulos, C. Brewster, and B. Espejo-Garcia, "Developing a Model for the Automated Identification and Extraction of Agricultural Terms from Unstructured Text," in *The 1st International Online Conference on Agriculture—Advances in Agricultural Science and Technology*, p. 94, MDPI, Feb. 2022.

[55] D. Tay, "Data Analytics and Programming for Linguistics Students: A SWOT and Survey Study," *Journal of Statistics and Data Science Education*, vol. 32, pp. 303–314, Sept. 2024.