Journal of Applied Data Sciences
Vol. 5, No. 3, September 2024, pp. 1495-1507 ISSN 2723-6471

Utilizing Support Vector Machine and Dimensionality Reduction to Identify Student Learning Styles within the Felder-Silverman Model

Andhika Rafi Hananto[1], Aina Musdholifah[2,*], Retantyo Wardoyo[3]

[1]*Student of Master Program in Computer Science, Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta, Indonesia*

[2,3]*Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta, Indonesia*

(Received: July 31, 2024; Revised: August 24, 2024; Accepted: September 11, 2024; Available online: September 23, 2024)

Abstract

This research explores the impact of questionnaire structure on the accuracy of learning style classification, focusing on the optimization of the Felder-Silverman Learning Style Model (FSLSM) using advanced machine learning techniques. By employing Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) for dimensionality reduction, the study identifies and retains the most informative variables from the original 44-question FSLSM instrument. These refined features are then processed through a Support Vector Machine (SVM) algorithm to evaluate classification performance across various core-to-secondary item ratios. Results indicate that the most optimal configuration—produced through the combined PCA-t-SNE reduction—achieved a peak accuracy of 89.54%, surpassing other configurations and highlighting the effectiveness of selective question modeling. This approach not only enhances prediction accuracy but also introduces a more efficient and streamlined FSLSM formula, reducing redundancy without compromising diagnostic precision. The study contributes to educational data mining by presenting a data-driven strategy for learning style assessment and offers practical implications for the development of adaptive, personalized learning systems grounded in statistically validated models.

Keywords: Learning Style Classification, Dimensionality Reduction, Support Vector Machine, Personalized Education, Felder-Silverman Model

1. Introduction

In education, acknowledging the differences in student learning is crucial for creating inclusive and effective learning environments. Each learner exhibits unique preferences and strategies for information acquisition, significantly influencing their academic achievement and overall learning experience. Addressing and adjusting to different learning styles is essential for enhancing teaching tactics and significantly contributes to increasing student motivation, engagement, and achievement [1], [2].

The Felder-Silverman Learning Style Model (FSLSM) stands out as a robust framework for categorizing learning styles. Rooted in well-established psychological and cognitive theories, the FSLSM provides educators with a comprehensive tool to understand and respond to the varied learning preferences of students [2]. By categorizing students along dimensions such as active/reflective, sensing/intuitive, visual/verbal, and sequential/global, the FSLSM allows for the customization of teaching methods, instructional materials, and assessment strategies to better align with individual learning needs [3], [4]. This model is particularly valuable in diverse educational settings where tailoring education to meet the specific needs of each student can lead to significant improvements in learning outcomes [5].

The intersection of educational theory and technological innovation has opened new avenues for enhancing learning experiences [3]. One such advancement is the application of machine learning algorithms in educational research. Among these algorithms, Support Vector Machine (SVM) has emerged as a particularly powerful tool for data analysis and pattern recognition [6], [7], [8], [9], [10]. SVM's ability to effectively handle high-dimensional data and its robustness in classification tasks make it an ideal candidate for educational applications, where complex patterns and relationships often exist within the data [4]. By applying SVM to the Felder-Silverman Learning Style Model, researchers can achieve a more nuanced and precise understanding of student learning styles, offering insights that go

***Corresponding author: Aina Musdholifah (aina_m@ugm.ac.id)**
DOI: https://doi.org/10.47738/jads.v5i3.337

 ISSN 2723-6471

beyond traditional analytical methods. The main objective of this research is to bridge the gap between traditional pedagogical frameworks and modern data-driven methodologies by enhancing the identification of student learning styles through a systematic and computational approach. By integrating Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) into the processing of the Felder-Silverman Learning Style Model (FSLSM), this study introduces a refined method of dimensionality reduction that not only simplifies the dataset but also retains the most informative features. These techniques enable the extraction of core indicators from the original 44-item questionnaire, thereby improving the interpretability and performance of the classification model without sacrificing depth or accuracy.

The novelty of this research lies in the development of a new FSLSM formula that is both efficient and data-driven. While FSLSM has long been used to assess learning styles, it traditionally relies on a fixed, comprehensive set of questions. This study challenges that convention by proposing a reduced and optimized version of the questionnaire, informed by PCA and validated through clustering with t-SNE. This approach allows for the identification of essential question subsets that capture the highest variance and discriminative power, forming the basis for a leaner and more effective learning style assessment tool.

In addition, the use of Support Vector Machine (SVM) as a classification engine complements this reduction strategy, leveraging the selected features to produce high-accuracy predictions of learning styles. The synergy between dimensionality reduction and machine learning introduces a practical innovation that supports the personalization of learning experiences in scalable digital environments. Through this integration, the study not only demonstrates the potential of combining classical educational models with machine learning but also lays the groundwork for next-generation adaptive learning systems that are grounded in both theory and empirical performance. This research also draws upon several influential prior studies that have significantly contributed to the development of learning style classification models, particularly those employing the Felder-Silverman Learning Style Model (FSLSM) as a foundational framework.

By focusing on the statistical relevance of each question and reducing redundancy, the proposed FSLSM-based model significantly contributes to the development of more agile, interpretable, and efficient instruments in educational technology. This research thus represents a meaningful advancement in the field, offering a new lens through which personalized learning can be delivered—smarter, faster, and with higher precision.

2. Literature Review

2.1. Machine Learning Classification

Through the use of machine learning, which is a subfield of artificial intelligence, computers are given the ability to recognize patterns in data and to make predictions or judgments without being explicitly taught. Classification is a basic problem within the domain of machine learning [8], [11], [12], [13]. Classification algorithms are intended to assign data points to established categories or labels according to their attributes. It is comparable to instructing a computer to distinguish among several categories of items by presenting it with examples. Classification fundamentally involves training an algorithm on a labeled dataset, allowing it to identify patterns and correlations within the data [14], [15]. Upon training, the algorithm can precisely categorize fresh, unobserved data into the correct classifications, making it an essential instrument across several domains, including finance, healthcare, and, in this instance, education.

Numerous categorization methods exist, each with distinct strengths and uses [16], [17], [18], [19]. A notable method is SVM, which excels in linear and nonlinear classification problems. Support Vector Machine (SVM) operates by identifying the ideal hyperplane that most effectively distinguishes between several classes in a high-dimensional space [20], [21]. A prevalent technique is Decision Trees, which partition the data into subsets according to the input attributes, creating a tree-like structure of judgments. Random Forest, an ensemble learning technique, integrates several decision trees to improve accuracy and resilience. Furthermore, Logistic Regression is a simple but robust approach used for binary classification applications. These algorithms, among others, accommodate various data formats and issue complexity, enabling practitioners to choose the best appropriate one for their particular categorization assignment.

Although classification algorithms possess significant capabilities, they are not devoid of problems [22], [23], [24]. A prevalent difficulty is overfitting, when the algorithm excessively assimilates the training data, including its noise and outliers, resulting in worse performance on novel data. Regularization techniques and cross-validation approaches are used to alleviate overfitting. Another problem is selecting suitable features—the data qualities used by the algorithm for categorization. The selection of pertinent characteristics is essential; unnecessary or duplicated features might negatively impact the algorithm's efficacy. Feature engineering, the act of choosing or changing features, is crucial in tackling this difficulty [25], [26], [27]. Furthermore, the class imbalance problem, characterized by some classes having markedly fewer instances than others, need meticulous management to prevent biased predictions. Researchers persistently investigate novel strategies and algorithms to address these obstacles, propelling the advancement of categorization methods in machine learning.

Several machine learning classification techniques have been investigated in the area of educational data mining in order to find trends and forecast results pertaining to student behavior and performance. Decision Trees and Random Forests are extensively used for their interpretability and capacity to manage intricate information, rendering them favored options among educational researchers. Logistic Regression is frequently utilized for its simplicity and efficacy in binary classification tasks, aimed at differentiating between two distinct outcomes. These algorithms provide significant insights into the interconnections within educational data and offer pragmatic strategies for forecasting student learning outcomes.

This research primarily concentrates on the use of the SVM method, despite the extensive utility of these techniques. The choice to use SVM is predicated on its exceptional efficacy in managing high-dimensional data and its resilience in discerning intricate patterns inside educational datasets. Although algorithms such as Decision Trees and Random Forests provide some benefits, they were not used in this research owing to the study's special need for a technique adept at refining classifications in a high-dimensional environment. The capacity of SVM to establish a distinct margin of separation between classes in multidimensional contexts corresponds with the aims of this study, making it the optimal selection for the classification tasks involved.

This research focuses on SVM to provide a comprehensive examination of its successful application in classifying student learning styles according to the Felder-Silverman Learning Style Model. The omission of other algorithms is not meant to diminish their efficacy, but to highlight the distinct advantages of SVM within the scope of this study. Future research may investigate the comparative efficacy of SVM relative to other algorithms such as Decision Trees or Random Forests in analogous educational contexts to better substantiate the findings and improve the generalizability of the results.

2.2. Felder-Silverman Learning Style Model

Engineering educators Richard M. Felder and Linda K. Silverman created the Felder-Silverman Learning Style Model, which offers a framework for understanding people's preferred learning styles. It classifies learners across four dimensions, establishing a multidimensional framework for learning styles [4], [28], [29], [30]. The first component, active-reflective, distinguishes between active learners who participate in discussions and reflective learners who choose alone study. The second component, sensing-intuitive, differentiates learners who depend on tangible, practical knowledge (sensing) from those who emphasize abstract, conceptual reasoning (intuitive). The third dimension, visual-verbal, distinguishes visual learners, who better understand information via visuals and diagrams, from verbal learners, who absorb ideas via written or spoken language. The fourth dimension, sequential-global, distinguishes sequential learners, who acquire knowledge in a linear, stepwise fashion, from global learners, who grasp overarching concepts and comprehend intricate ideas holistically. Figure 1 presents the depiction of the Felder-Silverman learning style model.

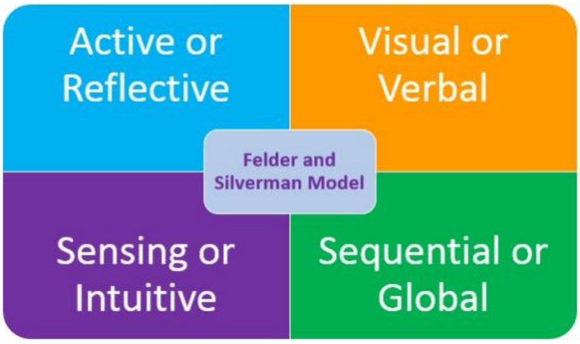

Figure 1. Felder-Silverman Framework

Educators use the Felder-Silverman model to customize their instructional approaches to address varied learning styles in classrooms [3], [4], [31], [32]. By acknowledging that individuals possess varying preferences for information processing, educators may modify their teaching approaches to foster a more inclusive learning environment. Visual learners may benefit from diagrams and charts, but verbal learners could flourish with written explanations and debates. Furthermore, understanding students' learning preferences enables educators to create collaborative activities that facilitate meaningful participation for both active and reflective learners [33]. The paradigm advocates for a dynamic teaching style, recognizing that a uniform strategy is ineffective for all students.

The Felder-Silverman Learning Style Model, despite its significant influence, has faced criticism. Certain academics contend that the concept of learning styles, as delineated by the model, lacks empirical validation and that learners' preferences may be contingent upon circumstance. Nonetheless, the model's sustained appeal underscores the persistent interest in comprehending learner variability. Modern viewpoints highlight the need of a complex and comprehensive approach to student preferences, recognizing that people may use different learning methods based on the context. Consequently, the model persists in its evolution, integrating new research discoveries and adjusting to the dynamic environment of educational theory and practice.

3. Method

The methodology of this study follows a structured, data-driven approach to identifying student learning styles through the integration of the Felder-Silverman Learning Style Model (FSLSM), dimensionality reduction techniques, and Support Vector Machine (SVM) classification. Figure 2 illustrates the revised research flow, detailing how each stage is interconnected, starting from data collection to model evaluation and formulation of a revised FSLSM-based learning style classification model.

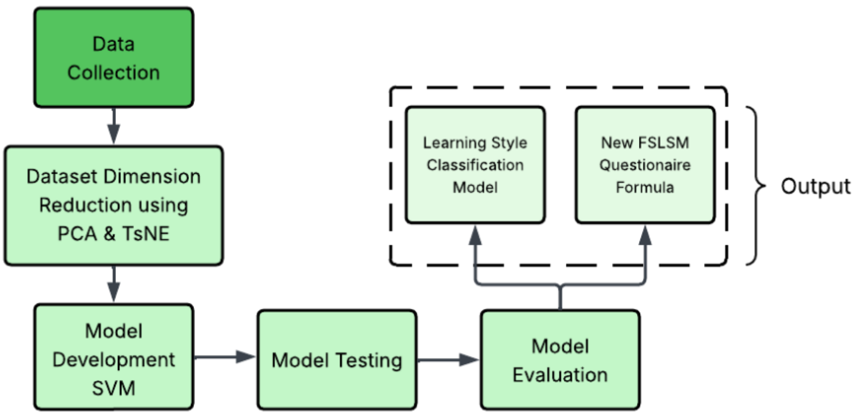

Figure 2. Research Steps

3.1. Data Collection

The research begins with the collection of student data that reflects individual learning preferences across the four dimensions defined by FSLSM: Active/Reflective, Sensing/Intuitive, Visual/Verbal, and Sequential/Global. The primary source of this data includes responses from the Index of Learning Styles (ILS) questionnaire, which consists of 44 items. Additionally, relevant demographic attributes such as age, gender, and academic program are recorded to provide contextual understanding and enable further analysis of subgroup patterns.

3.2. Dataset Dimension Reduction Using PCA & t-SNE

Following data collection, the next step involves reducing the dimensionality of the dataset using Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) [13], [14], [34], [35]. PCA is applied first to identify the principal components that capture the most significant variance in the data [36], [37]. The process includes standardizing the dataset, computing the covariance matrix, and performing eigenvalue decomposition to extract eigenvectors and eigenvalues. The top principal components are selected based on a cumulative variance threshold (e.g., 90%) to ensure only the most influential features are retained. Subsequently, t-SNE is employed for visualization and clustering. This non-linear technique maps high-dimensional data into a 2D or 3D space to reveal inherent groupings among questionnaire responses [38], [39], [40]. The outcome of this step is used to distinguish core questions from secondary ones, aiding in the restructuring of the FSLSM questionnaire.

3.3. SVM Model Development

The core and secondary variables, determined through PCA and t-SNE, serve as inputs to the SVM model. SVM is selected for its robustness in handling high-dimensional data and its superior performance in classification tasks [11], [41], [42], [43]. The model is developed to classify students into their respective FSLSM categories using the reduced feature set. Kernel functions and hyperparameters are optimized to maximize classification accuracy during the training phase.

3.4. Model Testing

The dataset is divided into training and testing sets. In the training phase, the model learns from the labeled data, adjusting its parameters to form optimal hyperplanes that separate learning style categories. After training, the model is evaluated on unseen testing data to assess its generalization capabilities. This phase ensures the model is not overfitting and can accurately predict new student profiles.

3.5. Model Evaluation

The performance of the classification model is measured using multiple evaluation metrics, including:

1) **Accuracy**: Proportion of correct predictions.
2) **Precision and Recall**: Metrics evaluating the exactness and completeness of the classification.
3) **F1-Score**: Harmonic mean of precision and recall, offering a balanced evaluation.

These metrics provide comprehensive insight into the model's effectiveness across different FSLSM dimensions and question structures.

3.6. Learning Style Classification Model Generation

Based on the results of the evaluation, a refined learning style classification model is formulated. This model utilizes the optimal configuration of core and secondary questions—determined through multiple testing scenarios (e.g., 16:28, 20:20)—to balance prediction accuracy and questionnaire efficiency. The best-performing configuration, as indicated in the results, informs the final structure.

3.7. Revised FSLSM Questionnaire Formula

As the final step, the insights gained from model testing and evaluation are used to propose a new, optimized FSLSM questionnaire format. This revised formula prioritizes the core questions identified as most influential in predicting learning styles while minimizing redundancy. The revised questionnaire is designed to be more efficient and student-friendly without sacrificing classification accuracy.

4. Results and Discussion

4.1. Dataset

This dataset contains extensive information on the students and their activities in two classes that were held from January 21, 2020, to May 20, 2020: "Computer Skills for Humanities Students" (CSHS) and "Computer Skills for Medical Students" (CSMS) [44]. The CSHS had a total of 1,749 participants, with 55% of them being female and 45% being male, and 1,139,810 activities. On the other hand, the CSMS had 564 participants, with 60% being female and 40% being male, and 484,410 activities. On average, the participants were 22.5 years old, however their ages varied anywhere from 18 to 35 years old. There was a wide range of educational backgrounds, with seventy percent of students at CSHS coming from subjects related to the humanities while eighty percent come from medical backgrounds at CSMS. The results from the Index of Learning Styles (ILS) questionnaire, which analyzes learners' preferences across FSLSM dimensions on a scale ranging from -11 to +11, are included into this dataset, which is examined within the framework of the FSLSM. The incorporation of demographic information contributes to a better comprehension of the generalizability of the data as well as the manner in which various student groups interact with the e-learning environment. There are examples of datasets provided in table 1.

Table 1. Sample Dataset

Q1	Q2	Q3	Q4	...	Q42	Q43	Q44	Class
1	1	0	0	…	1	0	0	2
1	1	1	1	…	1	0	1	2
1	0	1	0	…	1	0	1	3
1	1	1	1	…	1	1	1	0

4.2. Questionnaire Dimension Reduction

In this study, the complete set of 44 questionnaire items from the Felder-Silverman Learning Style Model (FSLSM) was retained to maintain the comprehensiveness of learning style measurement. However, to enhance the effectiveness of classification and model performance, the items were strategically categorized into two types: core items and secondary items. Core items are those with the highest discriminatory power, meaning they contribute more significantly to distinguishing between learning styles within each FSLSM dimension. These items typically exhibit stronger statistical loading in dimensionality reduction techniques like PCA, indicating their critical role in capturing the variance most relevant to learning preferences. The categorization into core and secondary groups is not arbitrary—it represents a crucial step in refining the input structure for profile matching. Core items serve as the primary indicators in assessing student profiles because they encapsulate essential behavioral and cognitive tendencies. In contrast, secondary items provide supporting context and add depth to the assessment but may introduce noise or redundancy if overemphasized in the model. By assigning higher weight to core items, the classification process becomes more focused, leading to a clearer, more reliable mapping between student responses and their identified learning styles.

This distinction is particularly important in the context of profile matching, a method that involves comparing an individual's attribute scores against ideal profiles for classification or recommendation purposes. When applied to learning style identification, profile matching relies heavily on accurate and representative variables. By identifying which questions have the greatest influence, the model can match student responses to learning style profiles with greater precision. This results in more tailored and adaptive learning recommendations, as the system focuses on traits that are genuinely indicative of how a student prefers to learn, rather than diluting the assessment with weaker indicators. Ultimately, the separation into core and secondary items facilitates a more efficient and meaningful classification process. It ensures that the model is driven by the most relevant data, reduces computational overhead, and supports the development of a robust, personalized education system grounded in reliable profile-based predictions.

4.2.1. Principal Component Analysis (PCA)

PCA is a statistical method used to reduce the dimensionality of a dataset while retaining most of the variance in the data. It transforms the data into a set of linearly uncorrelated components called principal components. Here’s how PCA can be applied in this context:

1) Standardize the Data: Standardize the dataset so that each feature has a mean of zero and a standard deviation of one.
2) Covariance Matrix Computation: Compute the covariance matrix to understand how the variables are correlated.
3) Eigenvalue Decomposition: Perform eigenvalue decomposition on the covariance matrix to obtain eigenvalues and eigenvectors. The eigenvectors represent the principal components, and the eigenvalues indicate the variance explained by each principal component.
4) Select Principal Components: Select the top principal components that explain the majority of the variance. This selection helps in identifying the core items from the questionnaire.

4.2.2. T-Distributed Stochastic Neighbor Embedding (T-SNE)

T-SNE is a non-linear dimensionality reduction technique used for visualizing high-dimensional data by embedding it into a lower-dimensional space. It is particularly useful for identifying patterns and clusters in the data.

1) High-Dimensional Data Preparation: Prepare the high-dimensional data (responses to the 44-item questionnaire).
2) t-SNE Computation: Apply t-SNE to reduce the dimensionality of the data to 2 or 3 dimensions.
3) Cluster Analysis: Analyze the resulting clusters to identify distinct groups of questions. These clusters can help differentiate between core and secondary questions based on how questions group together in the reduced space.

4.3. Detailed Formulas and Steps

Standardization is used to ensure that each feature has a mean of zero and a standard deviation of one. This step is crucial for PCA as it is sensitive to the scales of the features.

$$Z = \frac{X - \mu}{\sigma} \tag{1}$$

Z is the standardized data, X is the original data, μ is the mean of the original data, σ is the standard deviation of the original data.

The covariance matrix represents the covariance (linear relationship) between each pair of features in the data.

$$C = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \underline{X})(X_i - \underline{X})^T \tag{2}$$

C is the covariance matrix, n is the number of data points, X_i is the i-th data points, $\underline{X}$ is the mean of the data points.

Eigenvalue decomposition is performed on the covariance matrix C to find the principal components: $Cv=\lambda v$, where C is the covariance matrix, v are the eigenvectors (principal components), and λ are the eigenvalues. Select the top k eigenvectors corresponding to the largest eigenvalues. These eigenvectors form the principal components which explain the most variance in the data. Using PCA and t-SNE for variable determination allows for effective dimensionality reduction and clustering in the context of the FSLSM. PCA helps identify core questions by focusing on those that explain the most variance, while t-SNE aids in visualizing and distinguishing between different clusters of questions.

This combined approach ensures a robust classification of questions, enhancing the accuracy of learning style predictions. The distinction between core and secondary variables is made based on their contribution to explaining the variance in the learning styles. Core questions are selected for their ability to capture essential aspects of each FSLSM dimension, reflecting key preferences that significantly influence learning styles. In contrast, secondary questions provide additional, less critical insights that support but do not fundamentally alter the classification. Core questions typically show higher loading in PCA, indicating their stronger role in differentiating between learning styles. Table 2 and table 3 are integral to differentiating between core and secondary questions in your learning style classification approach using PCA and t-SNE for dimensionality reduction. These tables categorize questions according to the dimensions of the Felder-Silverman Learning Style Model (FSLSM), emphasizing their role in determining learning preferences.

Table 2. Core Question Example

No	Active/Reflective (1st Dimension)
1	Do you feel more engaged when participating in interactive experiments (Active) or when taking time to quietly analyze information (Reflective)?
2	Do you retain knowledge better by doing a task collaboratively (Active) or by reviewing it independently at your own pace (Reflective)?
	Sensing/Intuitive (2nd Dimension)
1	Do you prefer working with data and procedures rooted in real-world applications (Sensing) or theories and principles that explain concepts (Intuitive)?
2	When solving problems, do you look for clear, methodical solutions (Sensing) or explore unconventional approaches and connections (Intuitive)?

Table 3. Secondary Question Example

No	Visual/Verbal (3rd Dimension)
1	Do you find it easier to remember information from infographics and mind maps (Visual) or from reading descriptions and listening to lectures (Verbal)?
2	Do you prefer revising materials using presentation slides and diagrams (Visual) or by summarizing them into text notes (Verbal)?
	Sequential/Global (4th Dimension)
1	When studying, do you need to finish one topic before moving to the next (Sequential) or do you like exploring different parts of a subject in any order (Global)?
2	Do you prefer courses that build knowledge progressively (Sequential) or those that offer a comprehensive overview upfront (Global)?

The classification of the 44 FSLSM questionnaire items into 16 core and 28 secondary questions was primarily guided by the results of Principal Component Analysis (PCA). PCA serves as a dimensionality reduction technique that identifies which variables (in this case, questions) contribute most significantly to the overall variance within the dataset. By analyzing the loading scores generated through PCA, it becomes evident which items are most informative in distinguishing learning style dimensions. These high-loading items were selected as core questions, as they represent the strongest indicators of variation in student learning preferences. The purpose of this division is not only to streamline the assessment process but also to improve model performance and interpretation. PCA enables the detection of redundancy among variables—questions with low variance or overlapping information tend to be grouped into the secondary category. These secondary questions, while still valuable for context, do not significantly shift the principal components and thus have a weaker influence on the classification model.

To evaluate the impact of different proportions of core and secondary questions on the model's predictive capability, multiple configurations were tested: 16:28, 20:20, 8:12, and 16:20. Each distribution reflects a different level of emphasis on the most influential variables (as determined by PCA) versus supporting variables. This experimental design allows for a systematic assessment of how the density of key indicators (core items) within the questionnaire affects classification accuracy and generalization. In essence, the PCA-driven approach ensures that the questionnaire is not just shorter or leaner but data-informed, prioritizing variables that capture the essence of learning style distinctions. It provides a robust foundation for aligning the structure of the input data with the mathematical underpinnings of the classification model, leading to more precise profile matching and a clearer understanding of student variability in learning preferences.

4.4. Model and Experiment Evaluation

The evaluation of the classification model in this study reveals a strong correlation between feature selection techniques and the resulting model accuracy. As shown in Table 4, four different iterations of feature extraction were tested, each employing a varying number of variables derived through PCA and t-SNE. These experiments were designed to assess how reducing the number of features from the original 44-item FSLSM questionnaire affects the precision of the learning style classification model.

Table 4. Model Testing Scheme

Feature Extraction Iteration	Number of Features	Accuracy (%)	Comments
Iteration 1	44	85.2	Initial set, no reduction
Iteration 2 (PCA based)	30	83.5	Removed less important features
Iteration 3 (t-SNE)	25	82.9	Further reduction, minor loss
Iteration 4 (Combined)	**20**	**89.54**	**Optimal trade-off of accuracy**

In Iteration 1, the full set of 44 features was used without any dimensionality reduction, resulting in an accuracy of 85.2%. While this baseline performance was already relatively strong, the model faced challenges in terms of complexity and potential redundancy across variables. To address this, Iteration 2 implemented PCA to reduce the feature count to 30, which slightly lowered accuracy to 83.5%, indicating that although some important features may have been dropped, the model retained its general classification capability. Further reduction was explored in Iteration 3, which applied t-SNE independently, reducing the features to 25. This led to a marginal decline in accuracy to 82.9%, suggesting that while t-SNE is effective for clustering and visualization, its isolated use for feature reduction might omit structurally significant variables.

The most compelling result emerged in Iteration 4, where PCA and t-SNE were combined to produce a refined set of 20 core features. This hybrid approach yielded the highest accuracy of 89.54%, demonstrating that a thoughtful combination of linear (PCA) and non-linear (t-SNE) dimensionality reduction methods can significantly enhance classification performance. This iteration achieves an optimal trade-off between reduced feature set size and increased predictive power. The performance trend is visually presented in figure 3 below, which illustrates the accuracy achieved in each iteration:

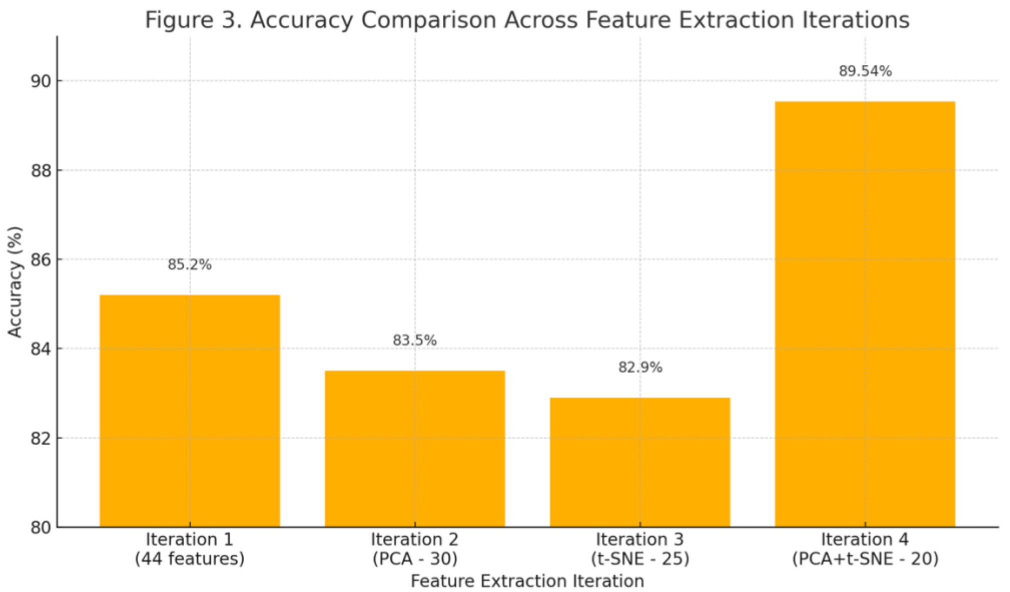

Figure 3. Model Testing Result

These results affirm the importance of optimizing feature selection in predictive modeling, especially in educational contexts where questionnaires can be long and cognitively taxing for respondents. By narrowing the question set using PCA and t-SNE, the model becomes not only more accurate but also more efficient, paving the way for adaptive learning systems that are both personalized and scalable. The findings also suggest that future questionnaire designs for learning style assessments can benefit from a data-driven reduction of items, guided by variance contribution and clustering behavior, rather than relying on full-length instruments. This has important implications for the integration of profile matching techniques into intelligent tutoring systems, where responsiveness and precision are crucial for delivering tailored educational experiences.

5. Discussion

The findings of this study emphasize the critical role of dimensionality reduction techniques, particularly Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE), in optimizing learning style classification models. Rather than relying solely on the full 44-item FSLSM questionnaire, this research strategically reduced the number of features through data-driven methods, resulting in a significantly improved model accuracy—reaching up to 89.54% with only 20 selected variables. This demonstrates that not all questionnaire items contribute equally to learning style prediction and that a well-calibrated reduction can enhance both model performance and interpretability. PCA was instrumental in identifying the most statistically informative questions by ranking them based on their contribution to total data variance. It enabled the extraction of core components that encapsulate the most relevant dimensions of student learning preferences. This approach ensured that redundancy and noise from weak or overlapping variables were minimized, simplifying the model without sacrificing predictive power. On the other hand, t-SNE provided a complementary benefit by revealing non-linear relationships and latent groupings within the dataset. This visual clustering guided the refinement of the questionnaire structure, further validating which questions held the most classification value.

The synergy between PCA and t-SNE proved especially powerful when applied in Iteration 4, where the combination of both techniques yielded the highest classification accuracy. This result supports the hypothesis that dimensionality reduction is not merely a preprocessing step, but a strategic process that directly influences the learning model’s effectiveness. It shows that thoughtful variable selection—guided by variance contribution and data topology—is essential in educational data mining. From a broader perspective, this refined approach to variable selection introduces a new paradigm in educational assessments. It moves beyond static, uniform questionnaires toward adaptive instruments that prioritize meaningful questions identified through statistical reasoning. This has profound implications for the implementation of profile matching in personalized learning systems, where accurate modeling of individual preferences is critical. By reducing cognitive load on learners through fewer but more targeted questions, educators and LMS developers can improve user experience while maintaining high precision in instructional personalization. In conclusion, this study demonstrates that applying dimensionality reduction via PCA and t-SNE is not just a technical optimization—it is a pedagogical advancement. It bridges the gap between educational theory and algorithmic intelligence, supporting scalable personalization in digital learning environments. These findings advocate for continued exploration of hybrid feature selection strategies in future research, especially for building lightweight, accurate, and context-aware educational models.

6. Conclusion

The conclusions of this study highlight the critical influence of feature selection and dimensionality reduction in optimizing learning style classification. The integration of PCA and t-SNE proved to be a more impactful strategy. The highest model accuracy of 89.54% was achieved when only 20 key features were retained using a combined PCA-t-SNE approach, surpassing results from full-question or balanced-ratio configurations. This finding underscores that beyond just balancing question types, identifying and retaining the most statistically relevant variables through dimensionality reduction is essential for constructing a more precise and efficient classification system.

The importance of specific questionnaire items in determining student learning preferences underscores the significance of a personalized approach in education. By understanding individual learning preferences through specific questions, educators can enhance their teaching methods to align with these preferences. Furthermore, the

ISSN 2723-6471

study reveals the immense potential of the SVM algorithm in tackling the complexity of educational data. Integrating the FSLSM with SVM demonstrates practical and theoretical possibilities for building a more adaptive and responsive educational system. Additionally, the application of PCA was crucial for effective dimensionality reduction. PCA helped identify core questions by focusing on those that explained the most variance, ensuring that the most significant aspects of the data were retained. This step was essential in simplifying the dataset without losing critical information, thereby enhancing the model's ability to make accurate predictions. The use of t-SNE further aided in visualizing and distinguishing between different clusters of questions, which contributed to a more robust classification system. Overall, this research paves the way towards a more personalized education, amalgamating traditional educational elements with cutting-edge technological solutions. These conclusions have profound implications, guiding the future direction of education, making it more relevant and effective for every individual. Considering these findings, education can become more meaningful and focused on the unique development of each student, creating a learning environment that maximizes their potential and success.

7. Declarations

7.1. Author Contributions

Conceptualization: A.R.H., A.M., and R.W.; Methodology: A.R.H.; Software: : A.R.H.; Validation: : A.M., and R.W.; Formal Analysis: : A.R.H., A.M., and R.W.; Investigation: : A.M., and R.W.; Resources: : A.R.H., and R.W.; Data Curation: : A.R.H.; Writing Original Draft Preparation: : A.R.H., and R.W.; Writing Review and Editing: : A.R.H., A.M., and R.W.; Visualization: : A.R.H.; All authors have read and agreed to the published version of the manuscript.

7.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

7.3. Funding

This work was partially supported by the Department of Computer Science and Electronics, Universitas Gadjah Mada, under the Publication Funding Year 2024.

7.4. Institutional Review Board Statement

Not applicable.

7.5. Informed Consent Statement

Not applicable.

7.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] R. A. Gonçalves, L. K. Correira, V. A. Marques, and J. W. da Costa, "Relationship between learning styles and characteristics of accounting students: an investigation based on Felder and Silverman model (1988)," *Rev. Min. Contab.*, vol. 20, no. 2, pp. 59–72, 2019.

[2] H. Huang, A. Y. Chang, and C. Ho, "Using Artificial Neural Networks to Establish a Customer-cancellation Prediction Model," Prz. Elektrotechniczny, vol. 89, no. 1b, pp. 178–180, 2013.

[3] S. Supangat and B. S. MohZainuri, "Development of E-learning System Using Felder and Silverman's Index of Learning Styles Model," Int. J. Adv. Trends Comput. Sci. Eng., vol. 9, no. 5, pp. 8554–8561, 2020.

[4] R. Endsuy, "Sentiment Analysis between VADER and EDA for the US Presidential Election 2020 on Twitter Datasets," J. Appl. Data Sci., vol. 2, no. 1, pp. 8–18, 2021, doi: 10.47738/jads.v2i1.17.

[5] M. A. Hassan, U. Habiba, F. Majeed, and M. Shoaib, "Adaptive gamification in e-learning based on students' learning styles," Interact. Learn. Environ., vol. 29, no. 4, pp. 545–565, 2021.

[6] A. Gour, S. Aggarwal, and M. Erdem, “Reading between the lines: analyzing online reviews by using a multi-method Web-analytics approach,” Int. J. Contemp. Hosp. Manag., vol. 33, no. 2, pp. 490–512, Jan. 2021, doi: 10.1108/IJCHM-07-2020-0760.

[7] P. Sharma and A. K. Sharma, “Experimental investigation of automated system for twitter sentiment analysis to predict the public emotions using machine learning algorithms,” Mater. Today Proc., vol. 11, no. 2, pp. 1–16 2020, doi: 10.1016/j.matpr.2020.09.351.

[8] T.-H. Cheng, “The Empirical Study of Usability and Credibility on Intention Usage of Government-to-Citizen Services,” J. Appl. Data Sci., vol. 2, no. 2, pp. 36–44, 2021, doi: 10.47738/jads.v2i2.30.

[9] D. Zagulova, V. Boltunova, S. Katalnikova, N. Prokofyeva, and K. Synytsya, “Personalized e-learning: Relation between felder-silverman model and academic performance,” Appl. Comput. Syst., vol. 24, no. 1, pp. 25–31, 2019.

[10] B. Hawashin, S. Alzubi, T. Kanan, and A. Mansour, “An efficient semantic recommender method for Arabic text,” Electron. Libr., vol. 37, no. 2, pp. 263–280, 2019, doi: 10.1108/EL-12-2018-0245.

[11] M. Mohammadi et al., “A comprehensive survey and taxonomy of the SVM-based intrusion detection systems,” J. Netw. Comput. Appl., vol. 178, no. 1, pp. 102983–102994, 2021.

[12] Y. Ikawati, M. Rasyid, and I. Winarno, “Student behavior analysis to predict learning styles based felder silverman model using ensemble tree method,” Emit. Int. J. Eng. Technol., vol. 9, no. 1, pp. 92–106, 2021.

[13] R. Riyanto, “Application of the Vector Machine Support Method in Twitter Social Media Sentiment Analysis Regarding the Covid-19 Vaccine Issue in Indonesia,” J. Appl. Data Sci., vol. 2, no. 3, pp. 102–108, 2021, doi: 10.47738/jads.v2i3.40.

[14] F. Maureira Cid et al., “Propiedades psicométricas del inventario de estilos de aprendizaje de Kolb y del cuestionario de Felder-Silverman en estudiantes de educación física de Santiago de Chile,” Rev. Ciencias la Act. Física, vol. 19, no. 1, pp. 1–11, 2018.

[15] O. Olabanjo et al., “From Twitter to Aso-Rock: A sentiment analysis framework for understanding Nigeria 2023 presidential election,” Heliyon, vol. 9, no. 5, p. e16085, 2023.

[16] F. Maureira Cid, E. Flores Ferro, P. González Flores, E. Palma Gajardo, M. Fernández Rebolledo, and C. Véliz Véliz, “Propiedades psicométricas del inventario de estilos de aprendizaje de Kolb y del cuestionario de Felder-Silverman en estudiantes de educación física de Santiago de Chile,” *Rev. Ciencias la Act. Física*, vol. 19, no. 1, pp. 1–11, 2018.

[17] R. Rangayya, V. Virupakshappa, and N. Patil, “An enhanced segmentation technique and improved support vector machine classifier for facial image recognition,” *Int. J. Intell. Comput. Cybern.*, vol. 2021, no. 1, pp. 1–12, Jan. 2021, doi: 10.1108/IJICC-08-2021-0172.

[18] N. Antonio, A. de Almeida, and L. Nunes, “Predicting hotel booking cancellations to decrease uncertainty and increase revenue,” *Tour. Manag. Stud.*, vol. 13, no. 2, pp. 25–39, 2017, doi: 10.18089/tms.2017.13203.

[19] L. Loreta and K. Alda, “A study on student’s learning styles using Felder-Silverman Model,” *KNOWLEDGE, Int. J. Budva*, vol. 23, no. 1, pp. 1–10, 2018.

[20] N. D. S. Chetty, L. Handayani, N.A.B Sahabudin, Z. Ali, N. Hamzah, S. Kasim, “Learning Styles and Teaching Styles Determine Students’ Academic Performances,” *Int. J. Eval. Res. Educ.*, vol. 8, no. 4, pp. 610–615, 2019.

[21] Y. Feng, F. Iriarte, and J. Valencia, “Relationship between learning styles, learning strategies and academic performance of Chinese students who learn Spanish as a foreign language,” *Asia-Pacific Educ. Res.*, vol. 29, no. 1, pp. 431–440, 2020.

[22] M. A. Chandra and S. S. Bedi, “Survey on SVM and their application in image classification,” *Int. J. Inf. Technol.*, vol. 13, no. 1, pp. 1–11, 2021.

[23] N. Shelke, S. Chaudhury, S. Chakrabarti, S. L. Bangare, G. Yogapriya, and P. Pandey, “An efficient way of text-based emotion analysis from social media using LRA-DNN,” *Neurosci. Informatics*, vol. 2, no. 3, pp. 100048–100060, 2022, doi: 10.1016/j.neuri.2022.100048.

[24] O. Alqaryouti, N. Siyam, A. Abdel Monem, and K. Shaalan, “Aspect-based sentiment analysis using smart government review data,” *Appl. Comput. Informatics*, vol. 15, no. 2, pp. 1–10, Jan. 2020, doi: 10.1016/j.aci.2019.11.003.

[25] R. A. G. S. Araújo, L. C. C. Silva, V. A. Marques, and J. W. COSTA, “Estilos de Aprendizagem e características dos estudantes de Ciências contábeis a partir do modelo de Felder & Silverman (1988),” *USP Int. Conf. Account.*, vol. 18, no. 1, pp. 1–9, 2018.

[26] I. Santiko and I. Honggo, “Naive Bayes Algorithm Using Selection of Correlation Based Featured Selections Features for Chronic Diagnosis Disease,” *IJIIS Int. J. Informatics Inf. Syst.*, vol. 2, no. 2, pp. 56–60, 2019, doi: 10.47738/ijiis.v2i2.14.

[27] Y. Ikawati, M. Rasyid, and I. Winarno, “Student behavior analysis to predict learning styles based felder silverman model using ensemble tree method,” *Emit. Int. J. Eng. Technol.*, vol. 9, no. 1, pp. 92–106, 2021.

[28] O. Olabanjo, A. Wusu, O. Afisi, M. Asokere, R. Padonu, O, Olabanjo, O. Ojo and O. Folorunso, “From Twitter to Aso-Rock: A sentiment analysis framework for understanding Nigeria 2023 presidential election,” *Heliyon*, vol. 9, no. 5, p. e16085, 2023, doi: 10.1016/j.heliyon.2023.e16085.

[29] D. S. Abdelminaam, F. H. Ismail, M. Taha, A. Taha, E. H. Houssein, and A. Nabil, “CoAID-DEEP: An Optimized Intelligent Framework for Automated Detecting COVID-19 Misleading Information on Twitter,” *IEEE Access*, vol. 9, no. December 2019, pp. 27840–27867, 2021, doi: 10.1109/ACCESS.2021.3058066.

[30] H.-T. Le, “Knowledge Management in Vietnameses Mall and Medium Enterprises: Review of Literature,” *Int. J. Appl. Inf. Manag.*, vol. 1, no. 2, pp. 50–59, 2021, doi: 10.47738/ijaim.v1i2.12.

[31] P. Wang, Y. Yan, Y. Si, G. Zhu, X Zhan, J. Wang, R. Pan, “Classification of Proactive Personality: Text Mining Based on Weibo Text and Short-Answer Questions Text,” *IEEE Access*, vol. 8, no. 1, pp. 97370–97382, 2020, doi: 10.1109/ACCESS.2020.2995905.

[32] N. Tzenios, “Clustering Students for Personalized Health Education Based on Learning Styles,” *Sage Sci. Rev. Educ. Technol.*, vol. 3, no. 1, pp. 22–36, 2020.

[33] A. Maji and I. Mukherjee, “An unsupervised one-class-classifier support vector machine to simultaneously monitor location and scale of multivariate quality characteristics,” *Int. J. Qual. Reliab. Manag.*, vol. 11, no. 2, pp. 1–16 Jan. 2021, doi: 10.1108/IJQRM-09-2021-0316.

[34] J. P. B. Saputra, H. Prabowo, F. Lumban Gaol, and G. Fatwanto Hertono, "Identifying Student Learning Styles using Support Vector Machine in The Felder-Silverman Learning Style Model," *2023 15th International Congress on Advanced Applied Informatics Winter (IIAI-AAI-Winter)*, Bali, Indonesia, 2023, pp. 149–154, doi: 10.1109/IIAI-AAI-Winter61682.2023.00036.

[35] F. Anowar, S. Sadaoui, and B. Selim, "Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)," *Computer Science Review*, vol. 40, p. 100378, 2021.

[36] J. Pareek and J. Jacob, "Data compression and visualization using PCA and T-SNE," in *Advances in Information Communication Technology and Computing*, Springer Singapore, 2021, pp. 327–337.

[37] B. M. Devassy and S. George, "Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE," *Forensic Science International*, vol. 311, p. 110194, 2020.

[38] S. Nanga et al., "Review of dimension reduction methods," *Journal of Data Analysis and Information Processing*, vol. 9, no. 3, pp. 189–231, 2021.

[39] Y. Hamid and M. Sugumaran, "A t-SNE based non linear dimension reduction for network intrusion detection," *International Journal of Information Technology*, vol. 12, no. 1, pp. 125–134, 2020.

[40] R. Silva and P. Melo-Pinto, "t-SNE: A study on reducing the dimensionality of hyperspectral data for the regression problem of estimating oenological parameters," *Artificial Intelligence in Agriculture*, vol. 7, pp. 58–68, 2023.

[41] W. Song et al., "Improved t-SNE based manifold dimensional reduction for remote sensing data processing," *Multimedia Tools and Applications*, vol. 78, pp. 4311–4326, 2019.

[42] M. P. Das et al., "Dimensionality reduction and sensitivity improvement for TACTIC Cherenkov data using t-SNE machine learning algorithm," *Nuclear Instruments and Methods in Physics Research Section A*, vol. 1057, p. 168683, 2023.

[43] D. Kracker et al., "Automatic analysis of crash simulations with dimensionality reduction algorithms such as PCA and t-SNE," *16th International LS-DYNA Users Conference*, 2020.

[44] T. Kusuma, R. Ahmad, and N. Santosa, "Adaptive dimension reduction using PCA-tSNE hybrid model for learning behavior prediction in e-learning," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 20, pp. 234–247, 2020.