G-Tech : Jurnal Teknologi Terapan
Vol. 10, No. 3, July 2026, p. 1469-1483
E-ISSN: 2623-064X | P-ISSN: 2580-8737

Identifying Financial Literacy and Asset Participation Segments among Young Adults in Indonesia Using K-Means Clustering

Oktavia Citra Resmi Rachmawati[1], Kevin Ilham Apriandy[2]✉, Kevin Harlis Oktaviano[3], Zakha Maisat Eka Darmawan[4]

[1, 2, 3] Program Studi D3 Teknologi Komputer, Politeknik Internasional Tamanasiswa Mojokerto, Indonesia
[4] Departemen Teknologi Multimedia Kreatif, Politeknik Elektronika Negeri Surabaya, Indonesia

✉ **Corresponding Author :** Kevin Ilham Apriandy (e-mail: kevinapriandy@gmail.com)

Article Information

Article History

Received : July 01, 2026
Revised : July 28, 2026
Published : July 30, 2026

***Keywords*:**

Financial Literacy;
Asset Participation;
Exploratory Data Analysis;
K-Means;
Clustering.

ABSTRACT

Financial literacy is crucial in influencing asset ownership decisions among young adults; yet, the variability of financial literacy and asset involvement in Indonesia has not been adequately examined. This research seeks to categorize young Indonesian individuals based on financial literacy and asset participation using the K-Means clustering technique. The research employed a quantitative methodology, incorporating exploratory data analysis of a survey dataset comprising 952 participants and 13 variables related to financial literacy, asset involvement, demographic traits, economic education, and financial behavior. Missing values were addressed by group-based mode imputation for categorical variables and mean imputation for numerical variables, followed by encoding and data standardization utilizing StandardScaler. The ideal number of clusters was assessed by the Elbow Method, Silhouette Score, and Davies–Bouldin Index. Despite achieving the highest Silhouette Score at k = 2, the k = 9 model was chosen due to its lower Davies–Bouldin Index and its ability to enable more nuanced responder segmentation. The findings identified nine categories exhibiting varying levels of basic and advanced financial literacy, ranging from very low to very high. These findings offer significant insights for the formulation of targeted financial education initiatives and financial inclusion policies customized to the attributes of various young adult demographics.

INTRODUCTION

Over the past two decades, Indonesia's financial system has become more complex, with more people gaining access to a variety of financial products and investment options (Suteja et al., 2023). Young adults are especially active in this shift, mainly because they are starting to manage their own finances and make independent decisions (Sinnewe & Nicholson, 2023). At this stage, they begin to form consumption habits, start saving and investing, and work toward owning assets that can affect their long-term financial well-being (Boto-García et al., 2022). However, better access to financial services does not always lead to stronger financial understanding (Panos & Wright, 2025). This can create a gap between what people know and how they actually manage their money (Pandeirot & Aseng, 2024). This is clear in Indonesia, where national financial literacy remains low while financial inclusion continues to rise, according to the Financial Services Authority (Rahayu et al., 2023).

This is an open access article under the CC BY license

Financial literacy helps people manage money and make informed financial decisions, including selecting assets that match their financial goals, risk tolerance, and long-term financial needs, while avoiding unnecessary financial risks (Lusardi & Messy, 2023). People with strong financial knowledge often know both the benefits and risks of different investments (Robb & Chy, 2023). These may include market assets such as stocks and mutual funds, and non-market assets such as savings, deposits, gold, and real estate (Oehler et al., 2023). People with lower levels of financial knowledge are more likely to own fewer assets, avoid risk, and prioritize short-term financial needs over long-term wealth accumulation, which may limit their opportunities to build financial security over time (Long & Tue, 2024). Research in Indonesia shows a strong link between financial literacy and money management (Sjam, 2025). However, existing studies have primarily examined the relationship between financial literacy and financial behavior using variable-centered approaches, providing limited insights into the heterogeneity of individuals with different financial literacy and asset participation profiles.

Young adults in Indonesia differ in their education, income, work experience, family background, and exposure to financial education. These variables were selected because previous studies have consistently identified them as important determinants of financial literacy and asset ownership, influencing individuals' financial knowledge, access to resources, and investment decisions. These differences can affect how much they know about finances and what assets they own. Segmenting people by similar traits is important for creating effective financial literacy policies and programs. Clustering methods like K-Means help group respondents with similar traits and reveal patterns in financial knowledge and asset ownership that simple analysis might miss.

Asset participation in early adulthood is shaped by more than just financial literacy, demographic traits, savings habits, economic education, and familial history. Family support and parental financial literacy have been associated with the development of saving and investment attitudes from an early age (Ndou & Godi, 2025). Also, lifestyle needs, consumption habits, and attitudes toward risk play a major role in asset ownership choices. To truly understand how financial literacy affects asset participation, it is important to study both the direct effects and the patterns within the early-adulthood group. This study aims to delineate divisions of early adulthood in Indonesia based on financial literacy levels and asset participation patterns using the K-Means clustering method. K-Means was selected because it is an efficient and widely used partition-based clustering technique for identifying relatively homogeneous groups within large datasets based on similarities across multiple variables (Ikotun et al., 2022), making it well suited to the objectives of this study. This methodology is anticipated to yield a more thorough understanding of the variability in early adult financial habits and to inform the creation of more successful and sustainable financial education and inclusion programs.

This study examines how financial literacy and asset participation patterns form among young adults in Indonesia. Prior research rarely segments early adulthood by analyzing the relationship between financial literacy levels and asset ownership using clustering methods. This study focuses on categorizing young adults in Indonesia based on their financial literacy and asset participation patterns. The study focuses on young adults in Indonesia who participated in the survey. Only respondents with complete and valid information on financial literacy, asset participation, and the demographic variables required for the clustering analysis were included in the study. This narrow scope helps maintain clarity. Financial literacy variables cover both basic and advanced knowledge, as well as self-evaluations of financial comprehension. Asset participation includes market assets (stocks and mutual funds) and non-market assets (savings, deposits, gold, and real estate). The sole analytical tool used was K-Means clustering to segment respondents, without exploring causal links among variables.

This study aims to identify and analyze the segmentation of young adults in Indonesia based on their financial literacy levels and asset participation patterns using the K-Means clustering approach. This study seeks to delineate the attributes of each segment to enhance comprehension of the diversity in financial behavior among young adults in Indonesia. The findings of this investigation are anticipated to yield both theoretical and practical advantages. This study aims to enhance the literature on financial literacy and asset participation by applying a clustering-based

segmentation approach, which remains relatively underexplored in the Indonesian context. The outcomes of this study are anticipated to inform the government, financial institutions, and educational entities in developing better-tailored financial literacy programs suited to the peculiarities of the young adult demographic. This study aims to enhance young adults' understanding of the significance of financial literacy in fostering asset ownership and promoting long-term financial well-being.

RESEARCH METHODS

This study employs a quantitative methodology, using unsupervised learning data analysis techniques, to segment young individuals in Indonesia based on their financial literacy levels and asset participation. This methodology was used since the research aim was not to examine causal links among variables, but to uncover latent patterns and variability in respondent traits that are not readily apparent through traditional statistical analysis.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \quad (1)$$

The K-Means clustering method served as the principal methodology for categorizing respondents with like features, facilitating a more thorough delineation of financial literacy and asset ownership profiles. The method partitions observations into K clusters by minimizing the within-cluster sum of squared distances between each observation and the centroid of its assigned cluster.

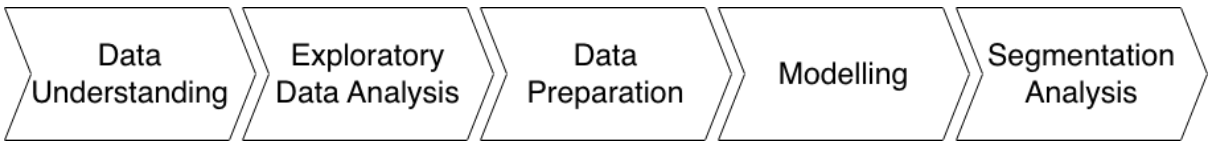


Figure 1. Research Methodology

The study utilized a secondary dataset obtained from the Mendeley Data repository (DOI: 10.17632/k269sxcw6h.1), which contains survey responses from 952 young adults in Indonesia collected to examine the relationship between financial literacy and asset participation. The dataset includes measures of basic, advanced, and self-assessed financial literacy; market asset participation (stocks and mutual funds); non-market asset ownership (savings, deposits, gold, and real estate); as well as demographic, behavioral, educational, and family background variables, including gender, academic year, employment status, income, savings rate, economics education, perceptions of economics in daily life, and parents' financial knowledge.The initial phase of this research involves data comprehension to gain preliminary insights into the dataset's attributes (Darmawan et al., 2025). At this juncture, descriptive statistical analysis is performed to elucidate the distributions of each research variable (Sari & Rachmawati, 2026), including basic financial literacy, advanced financial literacy, subjective financial literacy, market asset ownership, non-market asset ownership, and pertinent demographic and behavioral variables. Descriptive statistics, including the mean, median, standard deviation, minimum, and maximum, are used to summarize the patterns and variability in the respondent data (Rachmawati & Darmawan, 2024). A comprehensive assessment of missing values is performed to ensure the integrity of the data under analysis. Missing values are addressed using methods tailored to each variable's features, employing mean-based imputation for numeric variables or the most frequent category for categorical variables, thereby preserving the data's representativeness (Sari et al., 2026).

The second stage is exploratory data analysis, which visually examines data patterns, correlations, and distributions before the modeling process (Darmawan et al., 2026). Histograms were generated for all continuous variables in the dataset, including basic financial literacy, advanced financial literacy, subjective financial literacy, market asset ownership, and non-market asset ownership, to visually examine their distributions. Skewness values were then calculated to assess the degree of asymmetry, with values between -1 and +1 considered indicative of an approximately symmetric distribution, while values outside this range suggested moderate to

substantial skewness (Iacobucci et al., 2025). Outliers were identified using boxplots based on the interquartile range (IQR), where observations falling below Q1 - 1.5 × IQR or above Q3 + 1.5 × IQR were flagged as potential outliers (Phuttisen & Srisodaphol, 2024). The results of these analyses were used to evaluate the suitability of the data for K-Means clustering and to determine whether further preprocessing was required prior to clustering. Pie charts illustrate the distribution of asset ownership and respondents' demographic traits, thereby elucidating dominance or disparity within specific categories. Additionally, correlation heatmaps are employed to assess the strength of relationships among numerical variables, namely between financial literacy indicators and asset participation. This correlation study aims to uncover potential variable redundancy and verify that each variable provides pertinent information for the segmentation process, rather than to establish causal conclusions.

The third stage involves data preparation, which centers on modifying the data to align with the requirements of the K-Means clustering method. Specifically, missing values were handled, categorical variables were encoded into numerical representations where necessary, and numerical variables were standardized to ensure that all features contributed equally to the Euclidean distance calculation used by the K-Means algorithm. At this stage, categorical data, including gender, education level, and economic background, are converted to numeric values using one-hot encoding, as the K-Means algorithm requires numerical input and calculates similarities based on Euclidean distance. This transformation prevents artificial ordinal relationships between categorical variables and enables them to contribute appropriately to the clustering process. This approach ensures that categorical data can be mathematically analyzed by the clustering algorithm without introducing bias due to category order (Oktaviano et al., 2026). Subsequently, all numerical variables were standardized utilizing a standard scaler by centering each variable around its mean and scaling it according to its standard deviation. This process ensures that all variables are measured on a comparable scale with equal variance, preventing variables with larger numerical ranges from exerting a disproportionate influence on the Euclidean distance calculations used in the K-Means clustering algorithm. Standardization is essential, as the K-Means algorithm is particularly sensitive to differences in variable scale; without scaling, variables with a broader range of values may overshadow cluster development (Khan et al., 2025).

The fourth stage, modeling, is central to segmentation in this study. K-means clustering was used to group respondents based on similar financial literacy and asset involvement traits. The correct number of clusters was chosen by combining several assessment methods. The Elbow method was used to determine the optimal number of clusters by running the K-Means algorithm with different cluster sizes and calculating the within-cluster sum of squares (WCSS) for each solution. As the number of clusters increased, WCSS decreased because observations were assigned to more homogeneous groups. The optimal number of clusters was identified at the point where the decrease in WCSS became noticeably less pronounced, indicating that adding additional clusters provided only marginal improvements in cluster compactness. The Silhouette score measures cluster cohesion and separation, with higher values indicating better clustering (Mulyani et al., 2023). To further test clustering validity, the Davies Bouldin Index was calculated to compare the average within-cluster similarity with the separation between clusters. Lower DBI values indicate better clustering performance, as they reflect more compact clusters that are well separated from one another. The optimal number of clusters was selected by considering the Elbow method, Silhouette Score, and DBI together to ensure that the final clustering solution was both mathematically robust and practically meaningful.

The fifth stage, segmentation analysis, analyzes and ascribes meaning to the established clusters (Nur & Siregar, 2024). At this stage, analysts examine the attributes of each segment by contrasting the mean values of financial literacy measures, asset involvement levels, and demographic and behavioral factors within each cluster. This analysis informs the construction of profiles of early adulthood cohorts: those with high financial literacy and substantial asset involvement; those with intermediate literacy, characterized primarily by non-market assets; and those with low literacy and restricted asset ownership. This approach offers an overview of the diversity of financial behavior during early adulthood in Indonesia and forms a foundation for

developing policy implications and recommendations for specialized financial literacy programs tailored to each segment's features.

RESULTS AND DISCUSSION

The dataset includes multiple financial literacy metrics, including basic and advanced financial literacy, as well as respondents' self-assessed financial literacy. The data includes information on market asset involvement through stock and mutual fund ownership, as well as non-market asset ownership such as savings, deposits, gold, and property. This dataset is augmented with demographic variables including gender, academic year, and income level; behavioral variables indicating saving habits; educational variables representing economic educational background and perceptions of economics in daily life; and family background variables detailing parental financial knowledge. This dataset provides extensive micro-level evidence on the financial awareness and investment behavior of young adults in underdeveloped countries, making it pertinent for clustering-based segmentation analysis.

Table 1. Dataset Description

Feature	Type	Missing Value
Stock market participation	Nominal Categorical	191
Non-market asset participation	Nominal Categorical	0
Basic financial literacy index	Discrete Numerical	5
Advanced financial literacy index	Discrete Numerical	6
Self-assessed financial literacy	Ordinal Categorical	0
Gender	Nominal Categorical	18
Academic year	Ordinal Categorical	85
Employment	Nominal Categorical	113
Income	Ordinal Categorical	108
Savings rate	Ordinal Categorical	174
Economics education	Ordinal Categorical	9
Economics in daily life	Ordinal Categorical	8
Parents' financial knowledge	Ordinal Categorical	1

This study's dataset consists of 952 responses from young adults in Indonesia and 13 variables covering financial knowledge, types of asset ownership, and background details such as demographics, behavior, education, and family. Asset participation variables fall into two main groups: market asset participation and non-market asset involvement. Market asset participation shows whether someone owns assets such as stocks or mutual funds, and it has 191 missing entries. Non-market asset participation covers ownership of savings, deposits, gold, and property. It has no missing values, so it's fully usable in the study. Financial literacy is measured using three indicators: a basic financial literacy index, an advanced financial literacy index, and self-assessed financial literacy. The first two are numerical with 5 and 6 missing entries, respectively. Self-assessed financial literacy is an ordered categorical variable with no missing values and reflects respondents' judgments of their own financial knowledge.

This dataset includes various ancillary variables that enhance responder segmentation analysis in addition to the key variables. Demographic data include gender (18 missing values; nominal categorical) and academic year (85 missing values; ordinal categorical), indicating respondents' educational attainment. Individual economic conditions are indicated by employment and income variables. These are nominal and ordinal categorical with considerable missing values—specifically, 113 and 108 observations, respectively. This illustrates the diversity of respondents' working conditions and income levels. The savings rate variable is ordinal and has 174 missing values; it measures respondents' financial behavior and savings habits. The characteristics of the educational and familial environment are represented by variables such as economic education, economics in everyday life, and parental financial literacy. All of these are ordinal categorical with comparatively low missing values. The dataset's variable composition

exhibits a robust multidimensional nature. This facilitates segmentation analysis of early adulthood across financial literacy, asset participation, and socioeconomic and behavioral factors. The presence of missing values in certain variables necessitates meticulous handling during data preprocessing to prevent bias in analytical outcomes.

Table 2. Feature with Missing Values Imputation

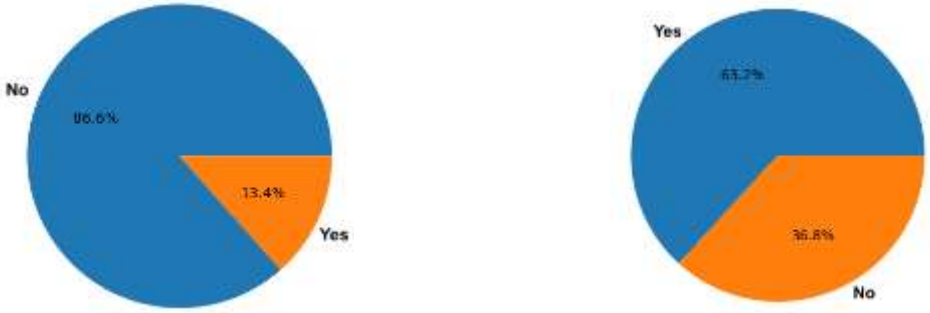
Feature with Missing Value	Dependent Feature	Statistic
Stock market participation	Non-market asset participation	Mode
Parents' financial knowledge	Self-assessed financial literacy	Mode
Academic year	Stock market participation	Mode
Gender	Academic year	Mode
Employment	Academic year	Mode
Economics in daily life	Self-assessed financial literacy	Mode
Economics education	Economics in daily life	Mode
Income	Employment	Mode
Savings rate	Academic year	Mode
Basic financial literacy index	-	Mean
Advanced financial literacy index	-	Mean

Missing values were addressed using a dependent-feature-based imputation method that fills in missing values by analyzing the interrelationships among variables, thereby preserving the original data structure. The imputation approach for categorical and ordinal variables is to use the mode value derived from grouping the pertinent dependent variable. For instance, missing values in the savings rate variable were imputed using the mode for each academic-year cohort, as illustrated in the data aggregation procedure. This enables the imputed values to accurately represent savings behavior inclinations at the same academic level. A comparable methodology was utilized for the stock market participation variable, which was inferred from non-market asset participation; parental financial acumen and economic understanding in daily life, which were deduced from self-reported financial literacy; and the academic year, gender, and employment variables, each of which was estimated using the mode value of the dependent variable exhibiting the most robust contextual correlation. The imputation procedure was systematic, accounting for the dataset's structure and logical relationships among variables.

	Stock market participation	Non-market asset participation	Basic financial literacy index	Advanced financial literacy index	Self-assessed financial literacy	Gender	Academic year	Employment	Income	Savings rate	Economics education	Economics in daily life	Parents' financial knowledge
0	0=otherwise	0=otherwise	-0.447789	0.849479	high	male	sophomores	not employed	Low (<IDR1,000,000)	NaN	some	a lot	intermediate
1	1=participate	0=otherwise	-1.388290	-0.576730	low	male	juniors	not employed	Low-Mid (IDR1,000,000-3,000,000)	NaN	a lot	NaN	low
2	0=otherwise	0=otherwise	0.233755	1.127099	low	female	sophomores	not employed	Low (<IDR1,000,000)	savings>consumption	little	a lot	high

Figure 2. Data Example

For the discrete numeric variables denoting the financial literacy index, specifically the basic and advanced financial literacy indices, the imputation approach employed was the mean value. This method was selected because both variables are quantitative and have relatively low rates of missing values, making mean imputation unlikely to substantially alter the data distribution. This study's missing-value management technique aims to preserve data pattern consistency and reduce potential bias in subsequent analyses, especially during the standardization and cluster-building phases. Integrating dependent-variable-based imputation for categorical data with descriptive-statistics-based imputation for numeric data preserves data quality, enabling the resultant segmentation to more accurately reflect respondent characteristics.



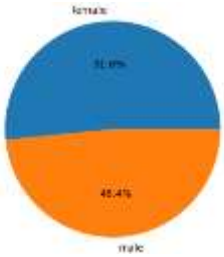
(a) Stock market participation (b) Non-market asset participation

Figure 3. The Distribution of Market

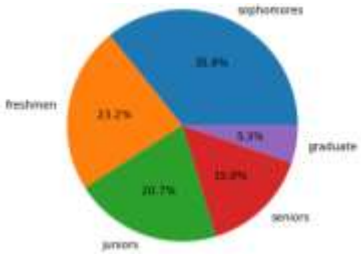
The pie chart shows a large difference in asset participation among young people in Indonesia. Only 13.4% invest in market assets such as stocks and mutual funds, while 86.6% do not. This highlights barriers such as financial knowledge, risk assessment, and limited access to market investments. In contrast, 63.2% own non-market assets such as savings, deposits, gold, or real estate, while 36.8% do not. This indicates that young adults prefer safer, more familiar, and accessible financial options over complex and risky ones, which is important for segmentation analysis.

(a) The Distribution of Savings Rate

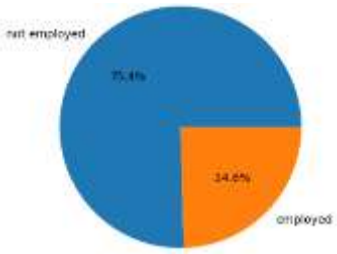
(b) The Distribution of Self-assessed Financial Literacy



(c) The Distribution of Gender

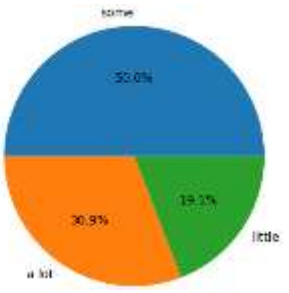


(d) The Distribution of Academic Year

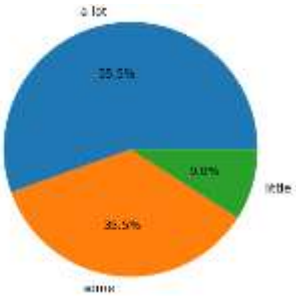


(e) The Distribution of Employment

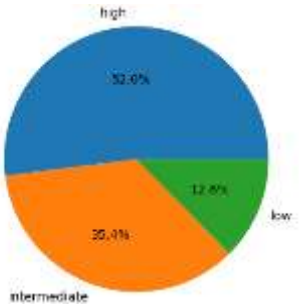
(f) The Distribution of Income



(f) The Distribution of Economic Education



(g) The Distribution of Economics in Daily Life

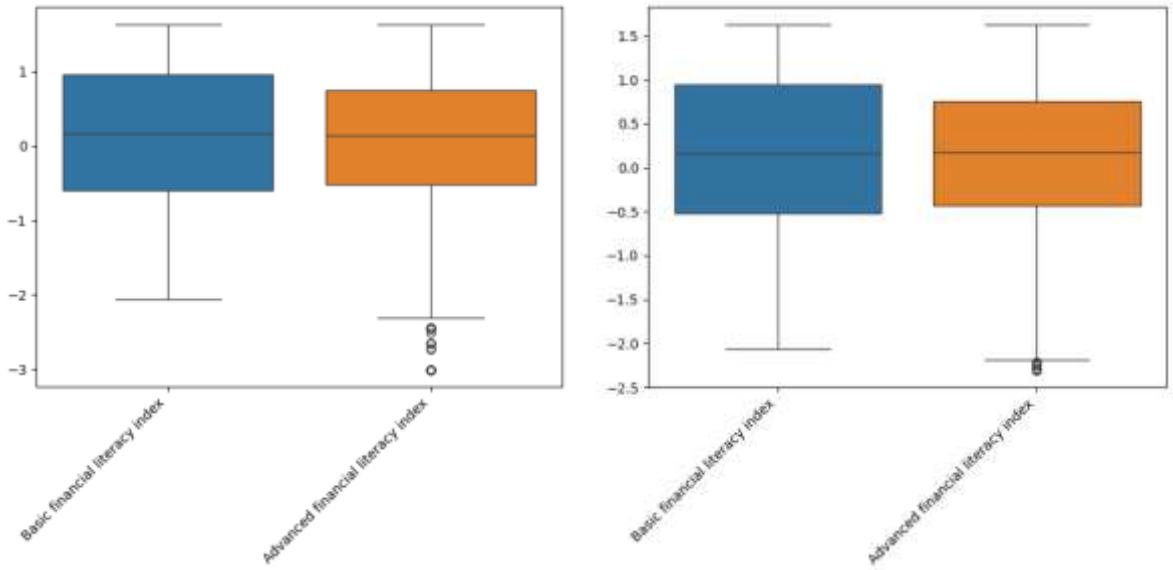


(h) The Distribution of Parents' Financial Knowledge

Figure 4. The Data Distribution of Features

This survey comprised 952 respondents, exhibiting a very equitable gender distribution of 51.6% female and 48.4% male. The predominant group of responses consisted of sophomores (35.8%), followed by freshmen (23.2%), juniors (20.7%), seniors (15.0%), and graduates (5.3%). The majority of respondents were unemployed (75.4%), whereas 24.6% were employed. This condition corresponds to the income distribution, indicating that 62.3% of respondents had low incomes (below IDR 1,000,000), while the remaining respondents fell into higher income brackets. Moreover, the data indicated that 73.3% of respondents had expenditures above their savings (savings < consumption), whereas approximately 13.1% managed to save more than they consumed. The data suggest that the majority of respondents remain in the nascent stages of financial independence, with limited ability to save.

The majority of respondents assessed their financial literacy abilities as poor (55.9%), followed by medium (28.3%) and high (15.9%). Nearly half of the respondents (50.0%) reported having received some form of economic education, while 30.9% acknowledged receiving substantial economic education. In their daily lives, 55.5% of respondents indicated that economics significantly influences their activities, suggesting a strong correlation between economic difficulties and their daily routines. Moreover, respondents generally viewed their parents' financial acumen favorably, with 52.0% assessing them as possessing high financial awareness and 35.4% as having intermediate financial understanding. The respondents are predominantly young adults who are primarily students or unemployed, have low incomes, and exhibit inadequate financial literacy, despite their exposure to economic education and a familial background with relatively sound financial comprehension.



(a) Before Outliers Removal (b) After Outliers Removal

Figure 5. Distribution of Variables Before and After Outlier Removal

During the data preparation phase, missing values were addressed by group-based mode imputation for categorical variables and mean imputation for numerical variables. This method entails imputing missing values by using categories with analogous attributes, yielding more pertinent information than merely using the dataset's general mode. For instance, missing values in the Stock Market Participation variable were imputed using the mode from the Non-Market Asset Participation group, whereas missing values in the Parents' Financial Knowledge variable were imputed using the mode from the Self-Assessed Financial Literacy group. A comparable procedure was implemented for the variables Academic Year, Gender, Employment, Economics Education, Income, and Savings Rate. For numeric variables, specifically the Basic Financial

Literacy Index and the Advanced Financial Literacy Index, missing values were replaced with the variable's mean to preserve the data distribution and avoid data omission.

After addressing all missing values, the subsequent phase involved converting categorical data to numeric using label and ordinal encoding. Binary variables, including Stock Market Participation, Non-Market Asset Participation, Gender, and Employment, are set to 0 or 1. Variables with defined levels, including Self-Assessed Financial Literacy, Income, Savings Rate, Economics Education, Economics in Daily Life, Parents' Financial Knowledge, and Academic Year, are transformed into an ordinal scale based on their respective levels. This transformation converts categorical variables into numerical representations, enabling them to be processed by the machine learning algorithms employed in this study.

Table 3. The Different Value between Before and After Data Standardization

Feature	Min		Max		Mean		Std	
	Before	After	Before	After	Before	After	Before	After
Respondent's identifier	1	-1,73023	952	1,73023	476,5	5,97094	274,96302	1,00052
Basic financial literacy index	-2,06433	-2,06542	1,632844	1,63370	-6,64202	1,50061	1	1,00052
Advanced financial literacy index	-3,00147	-3,00305	1,62403	1,62489	-2,35729	1,50220	0,99999	1,00052

The descriptive statistical comparison before and after the use of StandardScaler indicates that the standardization procedure effectively rescaled the data while preserving its original distributional pattern. Prior to standardization, each variable had a distinct range of values, characterized by the minimum (min), maximum (max), mean (average), and standard deviation (std), all of which differed among variables. Upon applying StandardScaler, all variables exhibited a mean near 0 and a standard deviation near 1, consistent with the aims of the standardization technique. Nonetheless, the minimum and maximum values of each variable were adjusted to a standard scale, thereby removing them from their original range. This modification indicates that StandardScaler applied a transformation by subtracting each value from the variable's mean and dividing by its standard deviation, thereby standardizing all variables to a uniform scale. Consequently, no single variable prevailed in the analysis due to disparities in units or value ranges, making the data more appropriate for use with diverse statistical and machine learning techniques that are sensitive to data scale (Garry et al., 2025).

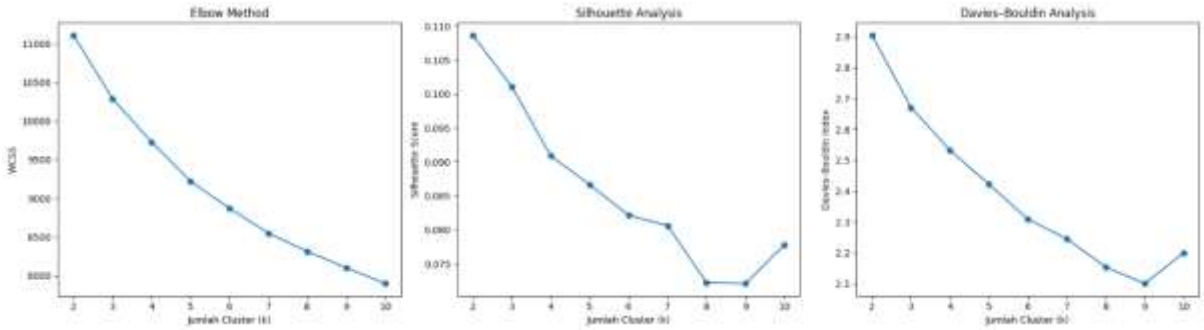


Figure 5. Model Evaluation Metrics

During the modeling phase, the K-Means Clustering algorithm was employed to categorize respondents according to financial characteristics, financial literacy, and asset

participation, following a pre-standardization procedure. K-Means was selected for its efficacy as a clustering technique for discerning groups of data with similar properties without requiring prior class labels (unsupervised learning). The modeling technique involved testing the number of clusters (k) ranging from 2 to 10. K-Means model was constructed for each value of k. A fixed random seed (random_state=42) was used to ensure reproducibility, allowing the clustering results to remain consistent across repeated runs. In addition, n_init=10 was specified so that the algorithm performs ten different centroid initializations and selects the solution with the lowest inertia, thereby reducing the likelihood of convergence to a suboptimal local minimum caused by unfavorable initial centroid placement. The clustering outcomes were subsequently assessed using three metrics: the Within-Cluster Sum of Squares (WCSS), the Silhouette Score, and the Davies–Bouldin Index (DBI). WCSS was used to measure cluster compactness and to identify the appropriate number of clusters using the elbow method. The Silhouette Score was employed to assess both intra-cluster cohesion and inter-cluster separation, with higher values indicating better-defined clusters. Meanwhile, the Davies–Bouldin Index evaluated the average similarity between clusters by considering both cluster compactness and separation, where lower values indicate better clustering performance. Using these three metrics provides a more comprehensive evaluation of clustering quality than relying on a single metric.

The initial assessment was performed with the Elbow Method with the WCSS value. The WCSS value reflects the overall variety of data inside a cluster; hence, a lower value signifies a more homogeneous cluster. The findings indicate a consistent reduction in the WCSS value from 11,109 at k=2 to 7,898 at k=10. This decrease occurs because increasing the number of clusters allows data points to be partitioned into smaller and more homogeneous groups, thereby reducing the distance between each data point and its corresponding cluster centroid. Consequently, the total within-cluster variation decreases as k increases. This decrease is a typical feature of K-Means, since increasing the number of clusters reduces the distances between data points and their respective cluster centroids. Nonetheless, the WCSS starts to stabilize after a certain number of clusters because additional clusters provide only marginal reductions in the within-cluster variation. At this stage, splitting existing clusters no longer produces substantial improvements in cluster compactness, indicating diminishing returns from increasing the number of clusters. Therefore, additional evaluation metrics, such as the Silhouette Score and the Davies–Bouldin Index, are required to determine the optimal number of clusters by considering both cluster compactness and cluster separation.

The cluster quality study revealed that the optimal Silhouette Score of 0.1087 was achieved at k=2, indicating a comparatively favorable separation between clusters compared with alternative values of k. Simultaneously, the minimum Davies–Bouldin Index was recorded at k=9, at 2.1008, signifying that the established clusters exhibit a commendable degree of intra-cluster similarity and distinctly defined inter-cluster separations. The disparity in outcomes between these two metrics indicates a compromise in selecting the number of clusters. The Silhouette Score favors simpler, more distinct clusters, whereas the Davies–Bouldin Index suggests that the data structure is better represented with a greater number of clusters. Consequently, determining the optimal number of clusters requires considering the evaluation metric values, the interpretability of the results, and the research objectives. In respondent segmentation, k = 2 is appropriate for achieving a straightforward, easily articulated segmentation, whereas k = 9 is suitable for obtaining a more nuanced segmentation to precisely identify respondent features.

Table 4. The Comparison of K Clusters Perfomance

k	WCSS	Silhouette Score	Davies–Bouldin Index
2	11109.370526	0.108651	2.904793
3	10289.184862	0.101102	2.670987
4	9730.447631	0.090909	2.532635
5	9221.093490	0.086652	2.422388
6	8873.677946	0.082139	2.309145
7	8549.553979	0.080630	2.246114

8	8313.372736	0.072263	2.153888
9	8103.355558	0.072103	2.100826
10	7897.841593	0.077810	2.200643

A comparison of clustering outcomes indicates that the k=2 cluster achieves a Silhouette Score of 0.108651, surpassing the k=9 cluster, which attains a score of just 0.072103. A superior Silhouette Score signifies that the constituents within each cluster at k=2 exhibit enhanced similarity and a more distinct distinction between clusters. The k=2 model can create two substantial, moderately compact, and distinctly identifiable clusters. The limited number of clusters amalgamates varied respondent characteristics into broader categories, potentially obscuring nuanced distinctions across respondent segments. Consequently, while k=2 yields superior separation quality, as indicated by the Silhouette Score, its segmentation outcomes are often overly simplistic to adequately capture the variability in respondent attributes.

In contrast, the k=9 cluster produces a Davies–Bouldin Index (DBI) of 2.100826, which is inferior to the k=2 cluster's 2.904793. A reduced DBI value indicates that the generated clusters exhibit greater internal homogeneity and more favorable inter-cluster distances. Despite the lower Silhouette Score at k=9, employing nine clusters facilitates the identification of more precise respondent segments by synthesizing financial literacy traits, income level, saving behavior, economic education, and asset participation. This analysis identified k=9 as the optimal number of clusters, as it offers a more nuanced and informative segmentation than k=2. This selection aligns with the aim of segmentation analysis, which is to elucidate variances in respondent characteristics in greater depth, enabling deeper analysis and interpretation of each group to facilitate decision-making and the formulation of policy recommendations.

Table 5. The Distribution of Clusters

Cluster	Count
0	91
1	96
2	100
3	119
4	94
5	64
6	154
7	115
8	119

The clustering results for k = 9 indicate that the distribution of members across clusters is rather uniform, with no significant predominance of any specific cluster. Cluster 6 comprises the largest cohort with 154 respondents (16.2%), followed by Cluster 8 and Cluster 3, each with 119 respondents (12.5%), and Cluster 7 with 115 respondents (12.1%). Additionally, Cluster 2 comprises 100 respondents (10.5%), Cluster 1 comprises 96 respondents (10.1%), Cluster 4 comprises 94 respondents (9.9%), and Cluster 0 comprises 91 respondents (9.6%). Cluster 5 comprises the fewest members, with 64 responses (6.7%). This distribution suggests that the resultant segmentation is notably balanced, as all clusters possess an adequate number of individuals for subsequent analysis. The absence of excessively small or large clusters indicates that the K-Means model effectively categorized respondents into numerous proportionate segments, enabling a more representative exploration and comparison of each cluster's attributes.

Tabel 1. Table Formatting (10pt)

Cluster	Basic Financial Literacy Index	Advanced Financial Literacy Index
0	-0.152148	-0.000785
1	-0.982135	-1.661444
2	0.031550	0.024149
3	-0.031662	0.366654
4	0.460936	-0.023512
5	0.770794	0.616412
6	-0.002420	0.242067
7	0.167720	0.402142
8	-0.023787	-0.060854

Each cluster demonstrates unique financial literacy traits based on the average scores of the Basic Financial Literacy Index and the Advanced Financial Literacy Index. Cluster 5 has the highest level of financial literacy, achieving the top average scores on both indices: 0.771 for basic literacy and 0.616 for advanced literacy. This indicates that members of this cluster possess a strong understanding of both fundamental financial concepts and complex financial instruments. In contrast, Cluster 1 has the lowest financial literacy, with average scores of -0.982 for basic literacy and -1.661 for advanced literacy. This cluster has scores significantly beneath the total respondent average, indicating a deficiency in financial knowledge across both fundamental and advanced dimensions.

The subsequent cluster exhibits greater trait diversity. Cluster 7 exhibits positive results on both indices (0.168 and 0.402), indicating medium to high financial literacy, particularly in advanced literacy. Cluster 4 exhibits a comparatively elevated basic literacy level (0.461), although its advanced literacy is nearly at the mean (-0.024), suggesting that individuals in this cluster possess a solid grasp of fundamental financial concepts but show lower proficiency in more complex financial tools. Clusters 3 and 6 show a notable trend: their basic literacy levels are near the average, while their advanced literacy levels are above average, at 0.367 and 0.242, respectively. This suggests that certain individuals within these groups possess a commendable understanding of specific financial items or instruments, despite their fundamental talents being quite unremarkable.

Clusters 0, 2, and 8 can be classified as possessing an average level of financial literacy. The results for both indices in these clusters are around zero, suggesting that their financial literacy traits are not markedly different from the overall average of all respondents. Cluster 2 exhibits a marginally positive value on both indices, making it slightly superior to the average, whilst Cluster 8 displays a marginally negative value on both indices, indicating a literacy level slightly below average. Cluster 0 is nearly identical to the mean, particularly with advanced literacy, which approaches zero. The clustering results indicate distinct segments, encompassing groups with minimal financial literacy (Cluster 1), average literacy levels (Clusters 0, 2, and 8), groups excelling in specific literacy aspects (Clusters 3, 4, and 6), and groups exhibiting the highest financial literacy (Clusters 5 and 7). This suggests that respondents possess a diverse range of financial comprehension levels, enabling the customization of financial education strategies to suit the characteristics of each section.

CONCLUSION

This research effectively delineated the segmentation of young adults in Indonesia based on financial literacy and asset participation using the K-Means clustering algorithm. The data preparation process, encompassing missing-value imputation, categorical variable encoding, and feature standardization, yielded a dataset suitable for clustering analysis. The evaluation employing the Elbow Method, Silhouette Score, and Davies–Bouldin Index revealed that these metrics may suggest different numbers of clusters. While the maximum Silhouette Score was

achieved at k = 2, the k = 9 configuration yielded the lowest Davies–Bouldin Index and provided more nuanced and significant segmentation. Consequently, the nine-cluster option was chosen as it more effectively represented the diversity of responders while ensuring equitable cluster sizes.

The segmentation results indicate significant variation in financial literacy among Indonesian young adults. Cluster 5 comprises respondents with the highest levels of both fundamental and advanced financial literacy, whereas Cluster 1 comprises respondents with the lowest levels of both. Multiple intermediate clusters display distinct traits, including comparatively robust basic literacy paired with average advanced literacy, or the reverse, suggesting that financial knowledge evolves in varying ways among responder groups. The generally even distribution of responses across the nine clusters demonstrates that the clustering algorithm effectively identified representative groups without producing disproportionately large or small clusters, thereby facilitating a more reliable interpretation of each segment.

This study has several limitations. First, the analysis was conducted using a single survey dataset of young adults in Indonesia, which may limit the generalizability of the findings to other populations or time periods. Second, the clustering results were based only on the variables available in the dataset and did not include additional behavioral or psychological factors that may influence financial literacy. Third, this study employed only the K-Means algorithm, and different clustering methods may produce different segmentation results. Furthermore, the clustering performance was evaluated using internal validation metrics without external validation.

Despite these limitations, the results of this study offer actionable insights for politicians, educational institutions, and financial service providers. Given that financial literacy traits vary across clusters, financial education initiatives should adopt a segmented strategy rather than a one-size-fits-all approach for all young adults. Individuals with limited financial literacy require essential financial education focused on budgeting, saving, and foundational investment principles, while those with greater literacy may benefit from advanced instruction on portfolio diversification, capital market engagement, and long-term financial strategy. Future research may address these limitations by incorporating additional behavioral and psychological variables, comparing different clustering algorithms, and validating the identified segments using external datasets to provide a more comprehensive understanding of financial behavior among young adults in Indonesia.

REFERENCES

Boto-García, D., Bucciol, A., & Manfrè, M. (2022). The role of financial socialization and self-control on saving habits. *Journal of Behavioral and Experimental Economics*, 100, 101903. https://doi.org/10.1016/j.socec.2022.101903

Darmawan, Z. M. E., Dianta, A. F., Fathoni, K., Rachmawati, O. C. R., & Apriandy, K. I. (2025). Comparison of Machine learning classification Methods for weather Prediction: A Performance analysis. *Jurnal Teknologi Terapan G-Tech*, 9(2), 715–727. https://doi.org/10.70609/gtech.v9i2.6649

Darmawan, Z. M. E., Rachmawati, O. C. R., Dianta, A. F., Fathoni, K., Hakkun, R. Y., Santoso, T. B., & Apriandy, K. I. (2026). Deep embedded clustering for Indonesian protein, fat, and energy availability data. *Journal of Applied Engineering and Technological Science (JAETS)*, 7(2), 1561–1579. https://doi.org/10.37385/jaets.v7i2.8996

Garry, E. M., Baglivo, A., Govil, P., Duryea, J. L., Liu, W., Lasky, T., Chakravarty, A., Rivera, D. R., & Bradley, M. C. (2025). Evaluating the impact of data standardization on Real-World data. *Pharmacoepidemiology and Drug Safety*, 34(8), e70191. https://doi.org/10.1002/pds.70191

Iacobucci, D., Román, S., Moon, S., & Rouziès, D. (2025). A tutorial on what to do with skewness, Kurtosis, and Outliers: new insights to help scholars conduct and defend their research. *Psychology and Marketing*, 42(5), 1398–1414. https://doi.org/10.1002/mar.22187

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2022). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

Khan, I. K., Daud, H., Zainuddin, N., & Sokkalingam, R. (2025). Standardizing reference data in gap statistic for selection optimal number of cluster in K-means algorithm. *Alexandria Engineering Journal*, 118, 246–260. https://doi.org/10.1016/j.aej.2025.01.034

Long, T. Q., & Tue, N. D. (2024). Financial Knowledge and Short- and Long-Term Financial Behavior Across Gender and Generations: Evidence from Japan. *SAGE Open*, 14(4). https://doi.org/10.1177/21582440241295846

Lusardi, A., & Messy, F. (2023). The importance of financial literacy and its impact on financial wellbeing. *Journal of Financial Literacy and Wellbeing*, 1(1), 1–11. https://doi.org/10.1017/flw.2023.8

Mulyani, H., Setiawan, R. A., & Fathi, H. (2023). Optimization of K value in clustering using silhouette Score (Case Study: Mall customers data). *Journal of Information Technology and Its Utilization*, 6(2), 45–50. https://doi.org/10.56873/jitu.6.2.5243

Ndou, A., & Godi, N. J. (2025). Moderating effects of parental income on the relationship between parental financial discussions and financial literacy of young adults. *International Journal of Social Economics*, 53(3), 503–515. https://doi.org/10.1108/ijse-07-2024-0622

Nur, M. F., & Siregar, A. (2024). Exploring the Use of Cluster Analysis in Market Segmentation for Targeted Advertising. *IAIC Transactions on Sustainable Digital Innovation (ITSDI),* 5(2), 158–168. https://doi.org/10.34306/itsdi.v5i2.665

Oehler, A., Horn, M., & Wendt, S. (2023). Investment in risky assets and participation in the financial market: does financial literacy matter? *International Review of Economics*, 71(1), 19–45. https://doi.org/10.1007/s12232-023-00432-9

Oktaviano, K. H., Apriandy, K. I., & Sunardiyanta, M. S. (2026). Design and implementation of a Web-Based Visual Search System for MSME E-Commerce using the FLASK Framework. *Jurnal Teknologi Terapan G-Tech*, 10(1), 371–380. https://doi.org/10.70609/g-tech.v10i1.8942

Pandeirot, L. B., & Aseng, A. C. (2024). Closing the Gap between Financial Literacy and Financial Inclusion. *Jurnal Informatika Ekonomi Bisnis*, 443–447. https://doi.org/10.37034/infeb.v6i2.877

Panos, G. A., & Wright, R. E. (2025). Financial teaching and financial understanding amongst young people. *Economics Letters*, 251, 112309. https://doi.org/10.1016/j.econlet.2025.112309

Phuttisen, S., & Srisodaphol, W. (2024). Detection of outliers method in grouped multivariate data based on the boxplot and Euclidean distance principles. *AIP Conference Proceedings*, 3158, 020021. https://doi.org/10.1063/5.0223817

Rachmawati, O. C. R., & Darmawan, Z. M. E. (2024). The Comparison of Deep Learning Models for Indonesian Political Hoax News Detection. *CommIT (Communication and Information Technology) Journal*, 18(2), 123–135. https://doi.org/10.21512/commit.v18i2.10929

Rahayu, F. S., Risman, A., Firdaus, I., & Haningsih, L. (2023). The Behavioral Finance of MSME in Indonesia: Financial literacy, financial Technology (Fintech), and financial attitudes. *International Journal of Digital Entrepreneurship and Business*, 4(2). https://doi.org/10.52238/ideb.v4i2.127

Robb, C. A., & Chy, S. (2023). Undergraduate financial knowledge, attitudes, and behaviors: The impact of financial life skills course on college students. *Financial Planning Review*, 6(1). https://doi.org/10.1002/cfp2.1155

Sari, S., & Rachmawati, O. C. R. (2026). Exploratory data analysis for monitoring the environment variables of sugarcane growth. *Optimum: Journal of Economics and Development (University of Ahmad Dahlan Yogyakarta).* https://doi.org/10.26555/jiteki.v11i4.31360

Sari, S., Rachmawati, O. C. R., & Sutikno, T. (2026). Adaptive sugarcane monitoring in Mojokerto using a hybrid powered IoT multi-sensor system and machine learning. *International Journal of Advances in Applied Sciences*, 15(1), 384. https://doi.org/10.11591/ijaas.v15.i1.pp384-395

Sinnewe, E., & Nicholson, G. (2023). Healthy financial habits in young adults: An exploratory study of the relationship between subjective financial literacy, engagement with finances, and financial decision-making. *Journal of Consumer Affairs*, 57(1), 564–592. https://doi.org/10.1111/joca.12512

Sjam, A. (2025). Survey Data on Financial Literacy and Asset Participation among Young Adults in Indonesia [Dataset]. In *Mendeley Data*. https://doi.org/10.17632/k269sxcw6h.1

Suteja, J., Gunardi, A., Alghifari, E. S., Susiadi, A. A., Yulianti, A. S., & Lestari, A. (2023). Investment Decision and Firm Value: Moderating Effects of Corporate social responsibility and profitability of Non-Financial Sector Companies on the Indonesia Stock Exchange. *Journal of Risk and Financial Management*, 16(1), 40. https://doi.org/10.3390/jrfm16010040