BAREKENG: Journal of Mathematics and Its Applications
September 2024 Volume 18 Issue 3 Page 1865–1878
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss3pp1865-1878

N-SOFT SETS ASSOCIATION RULE AND ITS APPLICATION FOR PROMOTION STRATEGY IN DISTANCE EDUCATION

Fatia Fatimah[1], Selly Anastassia Amellia Kharis[2*], Fauzan Ihza Fajar[3]

[1,2,3]*Department of Mathematics, Faculty of Science and Technology, Universitas Terbuka Pondok Cabe, South Tangerang, 15437, Indonesia*

*Corresponding author's e-mail: *selly@ecampus.ut.ac.id*

Article History:

Received: 21st February 2024
Revised: 11th April 2024
Accepted: 9th July 2024
Published: 1st September 2024

Keywords:

N-soft Sets;
Association Rule;
Decision-Making

ABSTRACT

In everyday life, we always encounter obstacles in seeing the interrelationships between several events to make the right decisions. Universitas Terbuka is a pioneer in distance education that implements digital transformation for new student registration, student services, and alums. The obstacle faced is determining a suitable promotion strategy for new students. As a result, a representative model is needed to handle such cases. As an extension of soft sets, N-soft sets can handle decision-making for binary and non-binary assessments. However, research has yet to be related to N-soft sets decision-making in data mining, especially association rule classification. This article proposes a new combination of N-soft sets with Association Rule (NSSAR). This article also introduces and applies the decision-making procedure using NSSAR to real. The population is new students of Universitas Terbuka Jakarta in the 2023/2024 odd semester. Samples were taken randomly using a questionnaire—primary data obtained by 201 new students. The following results are obtained based on the processed sample data using the NSSAR algorithm: 1) new students from Universitas Terbuka Jakarta are predominantly from Vocational High Schools domiciled in Bekasi, majoring in Bachelor of Management from the Faculty of Economics and Business; 2) The most favorite media information used by new UT Jakarta students is Instagram. Based on the results, the NSSAR algorithm gave relationship patterns between the number of new students based on region, study program, diploma of origin, and information media. Therefore, policymakers should consider the right promotional strategy to increase the number of students.

How to cite this article:

F. Fatimah, S. A. A. Kharis and F. I. Fajar., "N-SOFT SETS ASSOCIATION RULE AND ITS APPLICATION FOR PROMOTION STRATEGY IN DISTANCE EDUCATION," *BAREKENG: J. Math. & App.*, vol. 18, iss. 3, pp. 1865-1878, September, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Decision-making is an inevitable process in everyday life. Every individual, organization, or government must choose between options. Making effective decisions is required in both personal and professional contexts. Decision-making is becoming increasingly complex over time. Wrong decisions can have negative impacts both financially, reputationally, and socially. Technological developments, globalization, and social dynamics have become new perspectives that influence decision-making, so decisions that are taken require various approaches that are careful and beneficial. Decision-making has become critical in various fields, including computer science and data analysis. Data, which is increasingly abundant in the modern world, can be analyzed, and patterns can be considered in decision-making. One way to get patterns in decision-making is by using data mining. Several studies have used data mining to solve existing problems. Data mining is already used in health [1]-[3], education [4]-[6], economy [7], and other sectors.

One of the rapidly growing uncertainty theories is soft sets [8]. Soft sets theory is applied to coding [9], medical diagnosis [10], smartphone selection [11], human resources [12], and social network analysis [13]. A combination of soft sets is used in decision-making with positive and negative attributes [14], video recommendation systems [15], film selection [16], and association rules [17], [18]. Expansion of soft sets in the form of ratings called N-soft sets [19] can handle decision-making for binary and non-binary assessments. However, no research related to N-soft sets decision-making in data mining. Meanwhile, N-soft sets can help deal with the uncertainty and complexity of data mining caused by incomplete data, data noise, and many features. Therefore, research on N-soft sets decision-making in data mining has strategic value in optimizing performance, especially association rule. Soft sets are restricted to binary, but N-soft sets are connected to many-valued logic. The N-soft sets algorithm uses a threshold to represent binary and non-binary according to data or user needs.

Association Rule Mining (ARM) is a technique for identifying relationships and associations between items in a dataset. ARM provides a valuable picture of the relationship between elements, and this picture can be used as a consideration in making decisions. The concept of N-Soft sets and Association Rule Mining is expected to be a new idea as a powerful tool to help make more sophisticated and precise decisions. N-Soft sets refer to a generalization of classical set theory that allows a flexible combination of elements with different membership levels in a set. On the other hand, association rule mining involves identifying relationships and patterns in data that can help in decision-making, especially in big data analysis.

This research concerns mathematics innovation, especially developing N-soft sets combined with data mining. Data mining methods that can represent various types of data are needed. Combining N-soft sets with the association rule can represent various data types. In this point of convergence, the concept of N-Soft sets Decision Making appears in Association Rules. This concept combines the flexibility of the N-Soft set with the analytical power of ARM, providing a more comprehensive framework for addressing uncertainty and complexity in decision-making. Previous research conducted decision-making solely using N-Soft sets [20]-[22]. However, in this study, integrating N-Soft sets with ARM is introduced, representing a novel approach. This study aims to investigate the concept of N-Soft sets decision-making in association rules in more depth.

The resulting model is applied to the analysis of digital service users to strengthen the digital transformation of higher education in Indonesia. The challenges of digital services in higher education are related to users' acceptance across Indonesia. Universitas Terbuka is a pioneer in distance education that implements digital transformation for new student registration, student services, and alums. Determining suitable promotional media for new student recruitment is challenging. Therefore, this study aims to develop an N-Soft Sets Association Rule (NSSAR) decision-making algorithm and apply it to actual data to determine a suitable new student promotion strategy at Universitas Terbuka.

2. RESEARCH METHODS

This article is a theoretical and practical study. A theoretical study is needed to define the novelty related to N-soft sets with the Association Rule. So, we examine the underlying definition: soft sets, a combination of soft sets association rules, and N-soft sets. A decision-making algorithm N-soft sets association rule is created in the next stage. In practical studies, the algorithm findings are applied to real data, namely information media used by prospective new students. After the NSSAR decision-making algorithm has been created, it continues to apply to real data. The research population was new students at UT Jakarta for the 2023/2024 odd semester. Samples were taken randomly from the population, and students were asked to complete a questionnaire. Next, data cleaning was performed to ensure that no data was empty or duplicated. Then, the transformation was carried out using N-Softset and grouped using Association Rules Mining. The state of the art of the proposed new algorithm is presented in **Figure 1**.

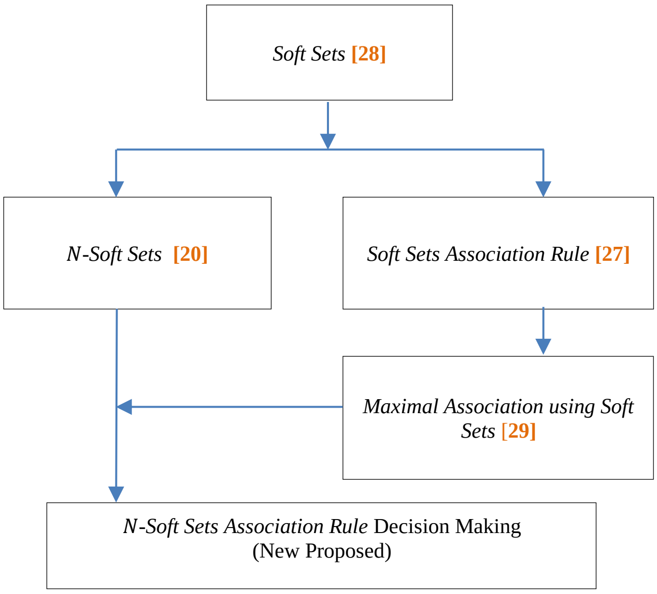

Figure 1. State of the art NSSAR

2.1 Basic Definitions

The notation U denotes the universe of objects, E symbolizes the universal set of parameters or attributes with $\boldsymbol{A} \subseteq \boldsymbol{E}$, and the set of ordered ranks is denoted by the notation $\boldsymbol{R} = \{\boldsymbol{0}, \boldsymbol{1}, \dots, \boldsymbol{N}-\boldsymbol{1}\}$, where $\boldsymbol{N} = \{\boldsymbol{2}, \boldsymbol{3}, \dots\}$.

In this section, definitions of related theories are presented. The definition of soft sets can be seen in **Definition 1**.

Definition 1 [8]. *Soft sets over the universe U, denoted as (F, A), are a mapping defined by the* **Equation (1)**.

$$F: A \to 2^U \tag{1}$$

Furthermore, the relationship of soft sets with association rules is elaborated in several definitions and algorithms. The definitions of softset can be seen in **Definitions 2 – 5**.

Definition 2 [18]. *Let (F, A) be a soft set over U and let $u \in U$. The set of co-occurrence items in transaction u is defined in* **Equation (2)**.

$$Coo(u) = \{a \in A: f(u, a) = 1\} \tag{2}$$

Can be written as $Coo(u) = \{a \in A: F(a) = 1\}$.

Interpretation of the F function is as follows. If u meets the criteria a, then it is written $F(a) = 1$. Otherwise, if u does not meet the property a, the value $F(a) = 0$ is given. Set A is usually interpreted as a list of characteristics, properties, or attributes (other interpretations are possible). Therefore, the domain of function F is denumerable sets.

Definition 3 [18]. *Suppose (F, A) is a soft set over U and two maximal itemsets $X, Y \subseteq E_i$, where $X \cap Y \neq \emptyset$. A maximal association rule between X dan Y is an implication of the form $Max\ (X \implies Y)$. Itemset X is the maximal antecedent, and Y is the maximal consequent.*

For decision-making using soft sets association, related definitions are required.

Definition 4 [18]. *Suppose (F, A) is a soft set over the universe U and two maximal itemsets $X, Y \subseteq E_i$, where $X \cap Y = \emptyset$. The maximal support of the maximal association rule $X \Rightarrow Y$, denoted as $Msup(Max\ (X \Rightarrow Y))$, is defined in* **Equation (3)**.

$$Msup(Max\ (X \Rightarrow Y)) = Msup(X \cup Y)$$

$$Msup(X \cup Y) = |\{u: X \cup Y = Coo(u) \cap E_i\}| \tag{3}$$

Definition 5 [18]. *Suppose (F, A) is a soft set over U and two maximal itemsets $X, Y \subseteq E_i$, where $X \cap Y = \emptyset$. The confidence of the maximal association rule, denoted as $Mconf(Max\ (X \Rightarrow Y))$, is defined in* **Equation (4)**:

$$Mconf(Max\ (X \Rightarrow Y)) = \frac{Msup(X \cup Y)}{Msup(X)}$$

$$\frac{Msup(X \cup Y)}{Msup(X)} = \frac{|\{u: X \cup Y = Coo(u) \cap E_i\}|}{|\{u: X = Coo(u) \cap E_i\}|} \tag{4}$$

Itemset X is referred to as the maximal antecedent, and Y is called the maximal consequent. Furthermore, it is an explanation of the definition of N-soft set.

Definition 6 [23]. *N-Soft Set (NSS) over U, denoted as (F, A, N), is defined as the following mapping in* **Equation (5)**.

$$F: A \rightarrow 2^{U \times R} \tag{5}$$

For each $a \in A$, there exist a unique ordered pair $(u, r_a) \in U \times R$ such that $(u, r_a) \in F(a) \in U, r_a \in R$.

2.2 Association Rules Mining (ARM)

Association rules mining is a data mining technique that discovers associative rules among item combinations. The concept of association rules involves finding all possible "if ⟶ then" relationships between items and selecting the most frequent item/itemset. According to Agrawal and Srikant, there are several statements about association rules. Let assume $I = \{i_1, i_2, \ldots, i_n\}$ is a set of items in transactions, and D is a set of transactions where each T is a set of items such that $T \subseteq I$. According to Agrawal and Srikant, the following is the definition of association rules, assuming X and Y are sets of items in I where $X \neq Y$. Association rules are expressed in the form of $X \rightarrow Y$, with $X \subset I, Y \subset I$, and $X \cap I = \emptyset$ [24].

Association rules use two parameters: support and confidence [25]. Support is the percentage of item combinations in a dataset. At the same time, confidence is the strength of the relationship between items in the association rule. One of the types of association rules mining is the apriori algorithm [26]. The apriori algorithm was introduced by Agrawal and Srikant in 1994 to find frequent items for Boolean association rules. Frequency pattern mining is one of the stages of association analysis that has garnered the attention of many researchers for generating efficient algorithms. The apriori algorithm aims to find frequent items when applied to a dataset. In each iteration k, all items with k items, known as k-itemset, are discovered. The first step in finding dataset association rules is finding frequent item sets. Frequent items are sets of items that frequently appear together. The main processes in the apriori algorithm to obtain frequent items are joining and pruning. Join involves combining items with other items until no more combinations can be formed, and prune involves trimming the results of the combined items using specified minimum support.

Support measures how frequently an itemset appears in the dataset as defines in **Equation (6)**.

$$support\ (A \rightarrow C) = \frac{\text{Number of transactions containing A and C}}{Total\ number\ of\ transaction} \tag{6}$$

Confidence measures the strength of the association rule, indicating the likelihood of item C being present in transactions that contain item A, as defined in **Equation (7)**.

$$confidence\ (A \rightarrow C) = \frac{support(A \cup C)}{support(A)} \tag{7}$$

Lift measures the ratio of the observed support of the itemset A∪C to the expected support if A and C were independent. It indicates how much more likely item C is to appear in transactions that contain item A compared to if C were independent of A. It is defined in **Equation (8).**

$$Lift\ (A \rightarrow C) = \frac{confidence(A \rightarrow C)}{support(C)} = \frac{support(A \cup C)}{support(A) X support(C)}. \quad (8)$$

2.3 N-SoftSets Association Rules (NSSAR)

The N-Soft sets Association Rules algorithm is a decision-making procedure used to discover association rules in a dataset using the concepts of the N-soft set. Algorithm N-Soft Sets Association Rules shown in **Algorithm 1**.

Algorithm 1. N-Soft Sets Association Rules

a. Suppose $U = \{u_i\}$ object subset, E is the set of parameters where $A = \{a\_j\} \subseteq E$ and $R = \{0,1,\dots,N-1\}$ denotes the rankings with $N = \{2,3,\dots\}$. Input a multi N-soft set(F, A, N) in such a way that for $\forall u_i \in U, a_j \in A, \exists! r_{ij} \in R$.

b. Specify the threshold T where $T \in R$ is the minimum limit.

c. If $r_{ij} \geq T$, then make changes $r_{ij} = 1$, others$r_{ij} = 0$.

d. Find the support of maximal association rules. Set the minimum support to 30%

e. Calculate the confidence of maximal association rules

f. Make decisions based on the relationship between objects and parameters determined by the highest confidence group

The implementation steps of the NSSAR algorithm can be seen in **Figure 2**.

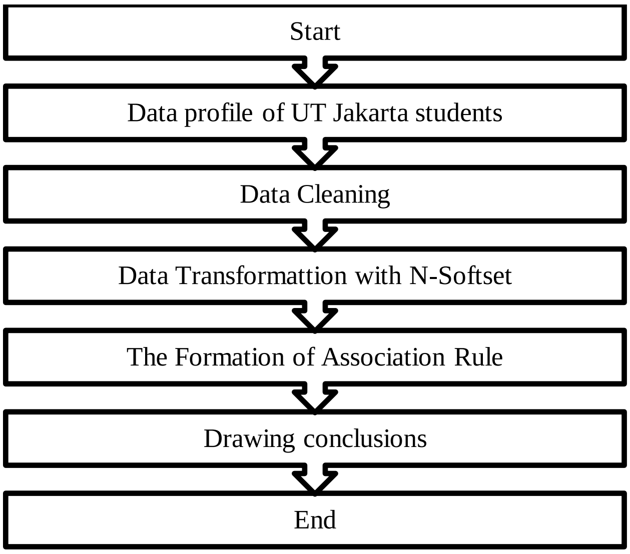

Figure 2. Steps of the NSSARM

Data was initially gathered by distributing questionnaires to new students at UT Jakarta. This stage involved collecting various demographic and academic information to create a comprehensive data profile. Following data collection, the data underwent a thorough cleaning process. This step was essential to remove any incomplete or duplicate entries. After cleaning, the data was transformed using the N-Softset methodology. This transformation was necessary to convert the raw data into a format suitable for association rule mining, facilitating the identification of meaningful patterns and relationships within the dataset. Using the transformed data, association rules were then formed. This process involved identifying sets of items or attributes that frequently co-occur within the dataset and uncovering hidden relationships among the various attributes of the students. Finally, conclusions were drawn based on the association rules formed.

3. RESULTS AND DISCUSSION

3.1 N-Soft Sets Association Rule Mining Application on New Students Data

The Distribution of new student admissions from the gathered data based on the origin school/institution before entering UT Jakarta is quite diverse. There are six categories of origin schools/institutions for the student data: Vocational High School, Senior High School, Islamic High School, Diploma Three, Package C, and transfer from Trisakti School of Management. Most origin schools/institutions come from Vocational High School, totaling 100 students.

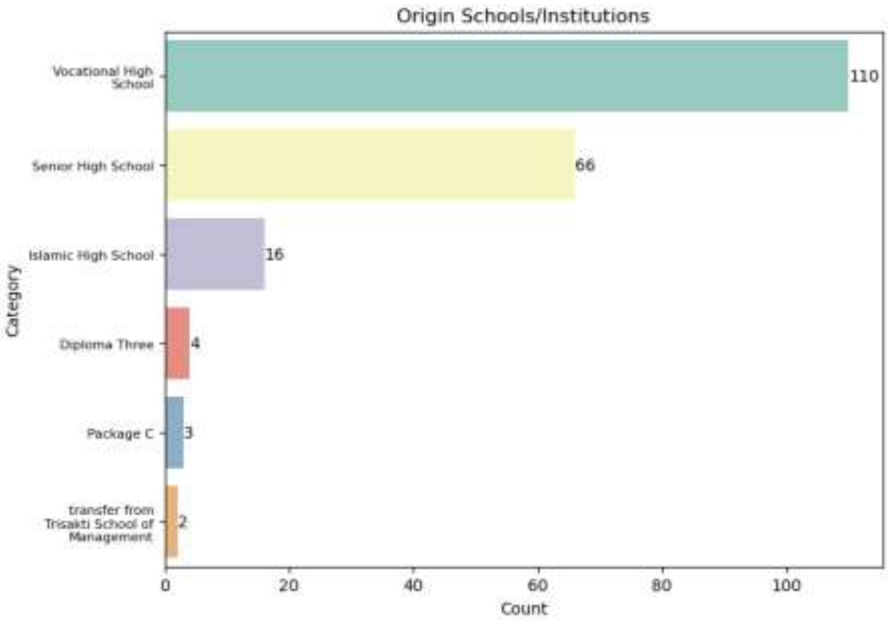

Figure 3. The distribution of students based on their origin schools/institutions

The detailed Distribution of students is shown in **Figure 3**. The student data is further analyzed based on the students' hometowns. In the data, several regional options have been merged; for example, Bekasi City and Regency are combined into the Bekasi region. Students who indicated Depok or Depok City are categorized under Depok. Likewise, Tangerang Regency and Tangerang Selatan are merged into Tangerang. The regions of Bogor City and Bogor Regency combined as Bogor. The regions Gunungkidul, Cirebon, Sukabumi, and Lampung are grouped as regions outside Jabodetabek—the Distribution of hometowns is shown in **Figure 4**, with most students coming from Bekasi.

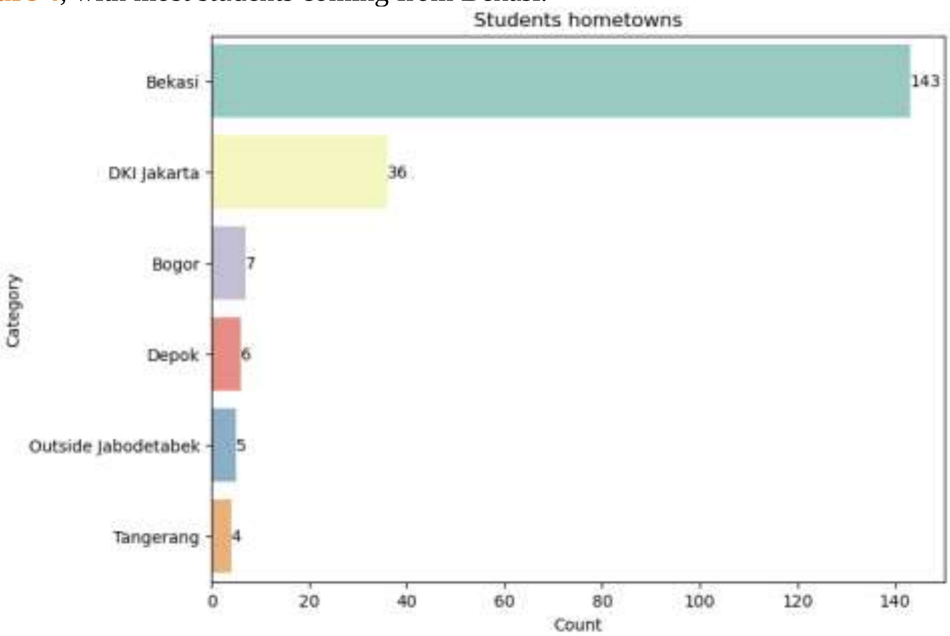

Figure 4. The distribution of students based on their hometowns.

The student data is further analyzed based on the Distribution of faculties and study programs. **Figure 5** shows the Distribution of student faculties, and **Figure 6** displays the Distribution of student study programs. Based on **Figure 5**, most students come from the Faculty of Economics and Business. It aligns with the Bachelor's program in Management, the most popular study program under the Faculty of Economics and Business.

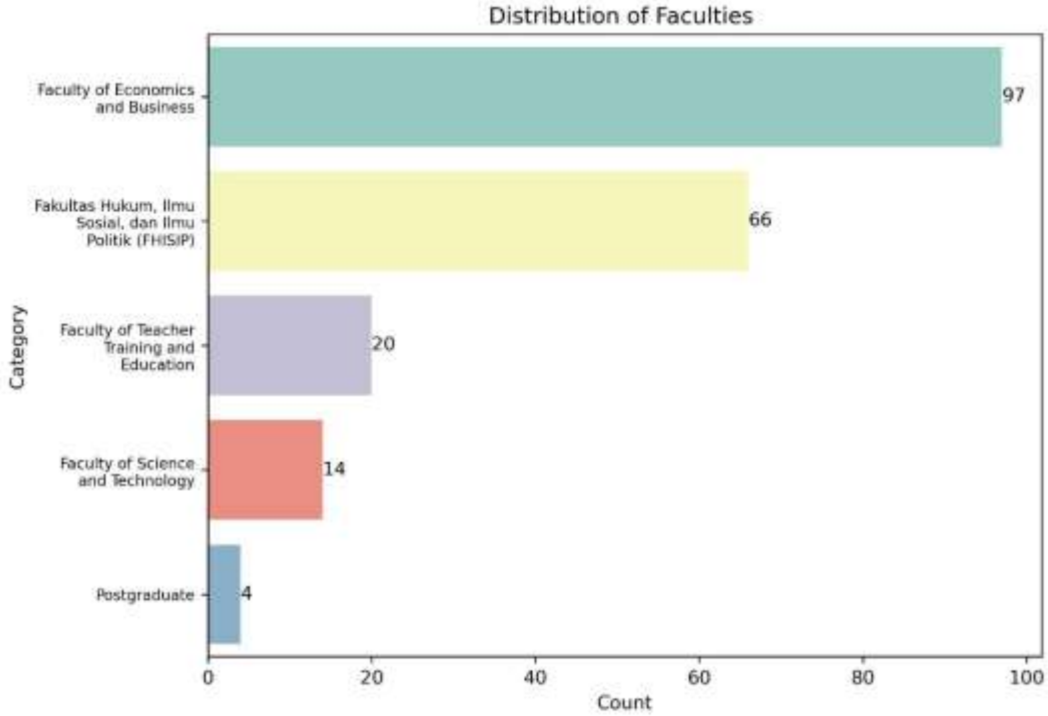

Figure 5. The Distribution of Students Based on Their Faculty.

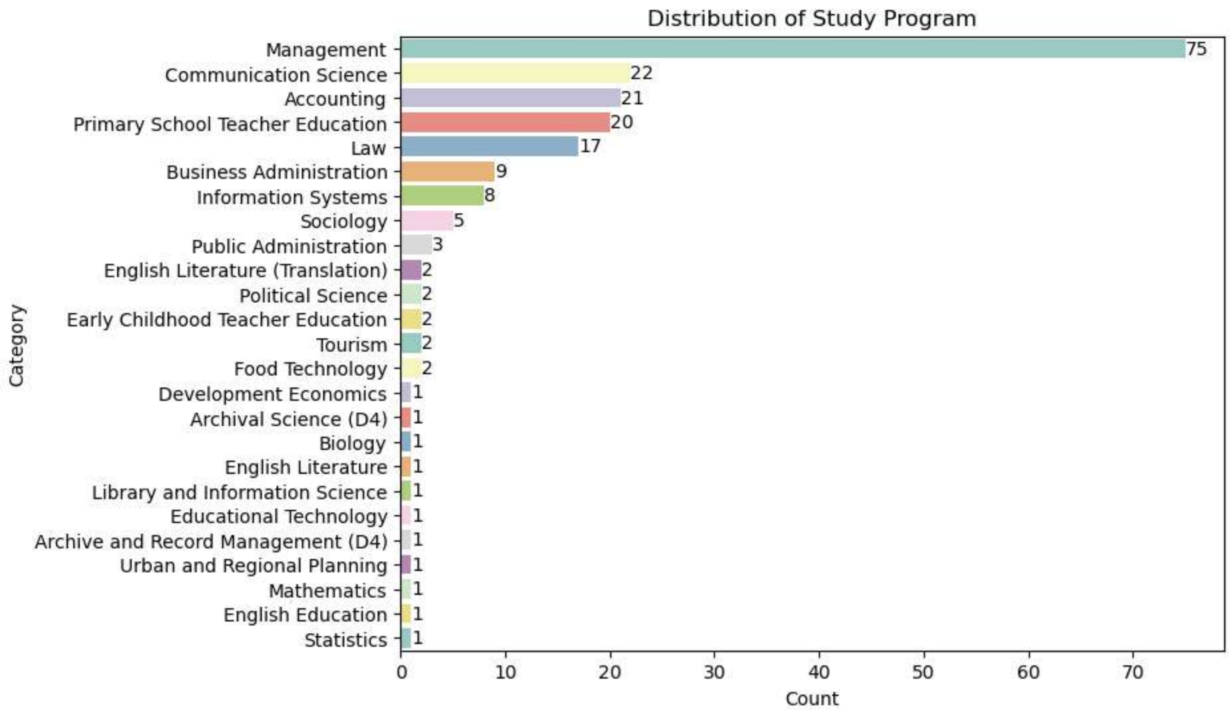

Figure 6. The Distribution of Students based on Their Study Program.

After understanding the existing student profile at UT Jakarta, preprocessing is initiated to tailor the data to the needs of the data mining process. During this stage, data selection is performed, focusing on relevant attributes. Selecting data in this manner aids in data mining by discovering valuable data patterns, constituting a transformation process. The data does not include attributes such as Student ID, Name, Phone Number, and the Type of courses taken at UT. The data used solely consists of the student's hometown, their previous school or institution before enrolling at UT Jakarta, their Faculty, Study Program, and the source of information about UT. The source of information about UT contains questions with a 5-point Likert scale. The Likert scale used comprises the following values: never (0), very rarely (1), sometimes (2), often (3), and always (4).

In this research, the data is grouped to facilitate the implementation of the previously planned system design under the required variables. The grouping will analyze the data essential to the system's design. The collected data cannot be processed directly; it needs to be transformed and assigned codes for ease of data processing. At this stage, the researcher groups the necessary data to determine the groups in the N-soft set. Students who choose a Likert scale of 0,1, or 2 are grouped into 0, while those who prefer 3 or 4 are grouped into 1. The researcher also performs checks for empty or duplicate data.

After preprocessing, the researcher proceeds to the N-Softset stage on the student data. N-Softset programming is implemented using Python. The N-Softset stage results provide binary data ready for grouping through association rules mining using Python. The source of information related to UT being studied includes promotional media such as Instagram, TikTok, Facebook, WhatsApp, study group leaders, friends/colleagues, family, billboards, banners, pamphlets, newspapers, magazines, radio, television, and websites. Patterns formed during the mining process are generated. Attribute testing is conducted to identify student groups with a minimum support of 30%. **Table 1** displays the formation of 15 groups with the highest confidence. The calculation of support using **Equation (6)** and confidence using **Equation (7)**

Table 1. The group with the highest 15 confidence

Antecedents	Consequents	Confidence	Support
The media sources are Instagram, vocational school, and non-family information sources	Information sources that are not Facebook, magazines, and billboards	1	0,31
The media sources are Instagram and information sources that are neither family nor billboards	Information sources that are not radio, Facebook, and magazines	1	0,31
The information sources are Instagram, vocational school, and non-family and non-magazine information sources	Information sources that are not radio, Facebook, and billboards	1	0,31
The information sources are Instagram, vocational school, and non-radio, non-family, and non-billboard information sources	Information sources that are not Facebook and magazines	1	0,31
The information sources are Instagram, vocational school, and non-magazine, non-family, and non-billboard information sources	Information sources that are not radio and Facebook	1	0,31
The information sources are Instagram, vocational school, and non-magazine, non-radio, and non-family sources	Information sources that are not Facebook and billboards	1	0,31
The information sources are Instagram, vocational school, non-Facebook, non-family, and non-billboard sources	Information sources that are not radio and magazine	1	0,31
The information sources are Instagram, vocational school, and non-Facebook, non-radio, and non-family sources	Information sources that are not magazines and billboards	1	0,31
The information sources are Instagram, vocational school, and non-Facebook, non-magazine, and non-family sources	Information sources that are not radio and billboards	1	0,31
The information sources are Instagram, vocational school, and sources that are not magazines, radio, family, and billboards	Information sources that are not Facebook	1	0,31
The information sources are Instagram, vocational school, and non-Facebook, non-radio, non-family, and non-billboard sources	Information sources that are not magazines	1	0,31
The information sources are Instagram, vocational school, and sources that are not Facebook, magazines, family, and billboards	Information sources that are not radio	1	0,31

Antecedents	Consequents	Confidence	Support
The information sources are Instagram, vocational school, and sources that are not Facebook, magazines, radio, and family	Information sources that are not billboards	1	0,31
The information sources are Instagram, vocational school, and sources that are not Facebook, family, and banners	Information sources that are not newspapers and magazines	1	0,31
The information sources are Instagram, vocational school, and sources that are not family and banners	Information sources that are not newspapers, Facebook, and magazines	1	0,31

The next step involves analyzing student data by grouping them based on their faculties. **Table 2** displays the results of the top 5 groups with the highest confidence, sorted by faculty. Next, an analysis of student data is conducted based on the student's hometown.

Table 2. The top 5 groups based on faculties with the highest confidence

Antecedents	Consequents	Confidence	Support
The information sources are Instagram, non-magazine information sources, radio, pamphlets, television, and banners, and the students are from the Faculty of Economics and Business	The information sources are not newspapers	1	0,30
The information sources are Instagram, non-newspaper and non-magazine information sources, pamphlets, television, and billboards, and the students are from the Faculty of Economics and Business	The information sources are not radio, Facebook, and banners	1	0,30
The information sources are Instagram, non-newspaper, non-magazine, pamphlets, television, and banners, and the students are from the Faculty of Economics and Business	The information sources are not radio, Facebook, and billboards	1	0,30
The information sources are Instagram, non-newspaper information sources, radio, pamphlets, television, and billboards, and the students are from the Faculty of Economics and Business	The information sources are not banners, Facebook, and magazine	1	0,30
The information sources are Instagram, non-newspaper information sources, radio, pamphlets, television, and banners, and the students are from the Faculty of Economics and Business	The information sources are not Facebook, magazines, and billboards	1	0,30

Table 3 displays the results of the top 5 groups with the highest confidence, sorted by students' hometowns.

Table 3. The top 5 groups based on students' hometowns with the highest confidence

Antecedents	Consequents	Confidence	Support
The information sources are Instagram, non-Facebook information sources, pamphlets, banners, and TikTok, and the student's hometown is Bekasi	The information sources are not magazines	1	0,30
The information sources are Instagram, non-pamphlet, and non-banners, and the student's hometown is Bekasi	The information sources are not Facebook and magazines	1	0,30
The information sources are Instagram, non-magazine, banners, and TikTok, and the student's hometown is Bekasi	The information sources are not Facebook and pamphlets	1	0,30
The information sources are Instagram, non-magazines, pamphlets, and TikTok, and the student's hometown is Bekasi	The information sources are not Facebook, and banners	1	0,30
The information sources are Instagram, non-Facebook, banners, and TikTok, and the student's hometown is Bekasi	The information sources are not magazines and pamphlets	1	0,30

A significant amount of data indicates a connection with information media, specifically Instagram. Therefore, the data was analyzed by considering students who use Instagram as antecedents. **Table 4** displays groups' results with antecedents of students using Instagram for information related to UT and their hometown in Bekasi. The calculation of lift using **Equation (8)**.

Table 4. The groups are based on the source of information related to ut using instagram and the students' hometown being in bekasi

Consequents	Confidence	Support	Lift
Information media that is not radio	0,958	0,453	1,02
Information media that is not a magazine	0,958	0,453	1,01
Information media that is not a newspaper	0,947	0,448	1,01
Information media that is not billboards	0,937	0,442	1,01
Information media that is not banners	0,937	0,443	1,02
Information media that is not Facebook	0,926	0,438	0,99
Information media that is not pamphlets	0,926	0,438	1,00
Information media that is not television	0,916	0,433	1,00
Information media that is not family	0,842	0,398	1,01
Information media that is not study groups	0,842	0,398	1,00
Information media that is not TikTok	0,653	0,308	0,97
Information media that is not newspapers, Facebook, magazines, family, pamphlets, study groups, television, billboards, and banners	0,653	0,308	0,95

Furthermore, the analysis is conducted on the results of groups with antecedents of students using Instagram for information related to UT and being from the Faculty of Economics and Business. The analysis results are shown in **Table 5**.

Table 5. The groups are based on the source of information related to ut using instagram and students from the faculty of economics and business

Consequents	Confidence	Support	Lift
Information media that is not banners	0,944	0,338	1,03
Information media that is not radio	0,944	0,338	1,00
Information media that is not newspapers	0,944	0,338	1,00
Information media that is not magazines	0,944	0,338	1,00
Information media that is not Facebook	0,944	0,338	1,00
Information media that is not billboards	0,931	0,333	1,00
Information media that is not pamphlets	0,903	0,323	0,98
Information media that is not television	0,875	0,313	0,95
Information media that is not study groups	0,875	0,313	1,03
Information media that is not family	0,847	0,303	1,00
Information media that is not newspapers, Facebook, magazines, radio, pamphlets, television, billboards, and banners	0,847	0,303	0,98

Table 6 shows the groups' results with antecedents of students using Instagram for information related to UT and coming from SMK. The analysis is then based on grouping solely by the antecedents of the students' hometown in Bekasi.

Table 6. The Groups based on the source of information related to UT using instagram and coming from SMK

Consequents	Confidence	Support	Lift
Information media that is not Facebook	0,899	0,353	0,96
Information media that is not banners	0,899	0,353	0,98
Information media that is not magazines	0,886	0,348	0,94
Information media that is not a newspaper	0,886	0,348	0,94
Information media that is not billboards	0,886	0,348	0,95
Information media that is not radio	0,873	0,343	0,93
Information media that is not pamphlets	0,861	0,338	0,93
Information media that is not television	0,848	0,333	0,92
Information media that is not family	0,810	0,318	0,98
Information media that is not newspapers, Facebook, magazines, radio, pamphlets, television, billboards, and banners	0,810	0,318	0,94

The grouping considered is for students who only use only one media. The results of this grouping are shown in **Table 7**. Based on the analysis of antecedents related to students' hometowns in Bekasi, students who rely on Instagram as an information source exhibit a confidence value of 66.4%, support value of 47.3%, and lift value of 0.98. For students who use websites as their information source, the confidence value is 46.9%, the support value stands at 33.3%, and the lift value is 0.98. Students who gather information from friends or colleagues have a confidence value of 46.9%, a support value of 33.3%, and a lift value of 0.92.

Table 7. The Groups based on Antecedents related to Students' Hometowns in Bekasi

Consequents	Confidence	Support	Lift
Information sources are Instagram	0,664	0,473	0,98
Information sources are website	0,469	0,333	0,98
Information sources are friends/colleagues	0,469	0,333	0,92

3.2 N-Soft Sets Association Rule Mining Data Analysis

Based on **Table 1**, the most prevalent group is formed when the antecedents consist of using Instagram as the information medium and coming from SMK. Students who use Instagram as an information medium and from SMK have various group variations. Although the antecedents and consequents are diverse, overall, students who use Instagram and come from SMK are not receiving UT-related information from their families, billboards, magazines, radio, Facebook, or banners. It suggests that decision-making regarding information media can be focused on Instagram for SMK students, as this group is the most common in the antecedents. Consequently, understanding these preferences can inform targeted strategies to enhancing UT-related communication channels for SMK students.

Regarding the students' faculties, the formed groups are exclusively from the Faculty of Economics and Business. It aligns with the majority of the data originating from students of the Faculty of Economics and Business. The majority of the formed groups use Instagram as the information medium. Additionally, based on **Table 3**, the groups are formed solely based on students' hometowns in Bekasi. Both grouping types yield a confidence value of 1 and a support value of 0.3. Confidence is one of the metrics used to analyse grouping results. In this context, a confidence value of 1 means that all association rules in this study have perfect confidence. In other words, every rule found is valid based on the data used. Meanwhile, a support value of 0.3 indicates that the association rules found are only relevant for items or elements that appear in at least 30% of the total observations in the dataset. This robust analysis underscores the reliability and applicability of the association rules derived, offering insights into effective promotional strategies tailored to specific student demographics and geographical locations.

The analysis using groups with antecedents of students using Instagram for UT-related information and coming from Bekasi shows that students who do not receive UT-related information from other media have a confidence above 60%. It suggests this demographic's strong reliance on Instagram for accessing UT-related information. Looking at the consequences, if UT wants to use Instagram as the sole information medium for marketing and target students in Bekasi, without using other media such as newspapers, Facebook, magazines, radio, family, pamphlets, study group leaders, television, billboards, and banners, this can be achieved with a confidence level of 65.3%. This finding underscores the effectiveness of Instagram as a singular promotional platform in reaching and engaging students in Bekasi. Furthermore, this analysis provides insights into the varying levels of confidence associated with different promotional media, including Instagram, radio, magazines, newspapers, billboards, banners, Facebook, pamphlets, television, family, study group leaders, TikTok, or combinations of these media. Understanding these confidence levels can guide strategic decisions in optimizing promotional efforts tailored to specific student demographics and media preferences.

The further analysis focuses on Instagram as the information medium for students from the Faculty of Economics and Business. Most students who use Instagram from the Faculty of Economics and Business without receiving information from other media have a confidence above 80%. It concludes that future marketing efforts for prospective Faculty of Economics and Business students can be focused on Instagram. The groups based on the school of origin that appear in NSS-ARM are only SMK. When Instagram is added as an antecedent, it leads to a conclusion that is almost similar to the groups based on the origin of students' place of residence and faculty. Based on **Table 7**, it is evident that if UT wants to promote to prospective students living in Bekasi, promotion can be carried out by disseminating information through Instagram, with a confidence level of 66.4%, through a website with a confidence level of 46.9%, and through friends/colleagues with a confidence level of 46.9%. These findings from the N-Soft Sets Association Rule Mining provide insights into the strategic use of Instagram and other media for targeted marketing efforts aimed at students from specific educational backgrounds and geographical locations, particularly emphasizing the role of Instagram in influencing decision-making behaviors among SMK students and those from the Faculty of Economics and Business.

4. CONCLUSIONS

N-soft set provides a solution for N-array assessment in soft set theory. This article introduces an innovation, which is the application of the N-soft set association rule for decision-making. The algorithm is applied to real data on new students at Universitas Terbuka. Based on the N-soft set association rule algorithm, it is revealed that most Universitas Terbuka Jakarta students come from Vocational High Schools, are domiciled in the Bekasi area, and enroll in the undergraduate management program at the Faculty of Economics and Business. Most of the new UT Jakarta students use socialization and information sources from Instagram. The implications and contributions of this research are to produce novel decision-making using N-soft sets on association rules to obtain representative, effective, and efficient algorithms to support the performance of the data mining process. For future research, the author intends to explore the novelty of combining N-soft sets with artificial bee colonies. In forthcoming research, it is also important to seek solutions for applying NNS decision-making to large datasets.

ACKNOWLEDGMENT

The authors express gratitude to The Directorate General of Higher Education, Research and Technology of the Republic of Indonesia for providing research funding for this article in the form of a research grant in the year 2023.

REFERENCES

[1] Z. Rustam and S. A. A. Kharis, "Comparison of Support Vector Machine Recursive Feature Elimination and Kernel Function as feature selection using Support Vector Machine for lung cancer classification," in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Jan. 2020. doi: 10.1088/1742-6596/1442/1/012027.

[2] S. A. A. Kharis and A. H. A. Zili, "Predicting life expectancy of lung cancer patients after thoracic surgery using SMOTE and machine learning approaches," *Jurnal Natural*, vol. 23, no. 3, pp. 152–161, 2023, doi: 10.24815/jn.v23i3.29144.

[3] S. A. A. Kharis, A. I. Tarigan, and D. Idayani, "Classification of lung cancer using support vector machine with feature selection based on artificial bee colony rate of change," in *International Seminar on Mathematics, Science, and Computer Science Education (MSCEIS)*, 2023, p. 080011. doi: 10.1063/5.0156155.

[4] S. A. A. Kharis *et al.*, "Design of Student Success Prediction Application In Online Learning Using Fuzzy-Knn," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 2, pp. 0969–0978, Jun. 2023, doi: 10.30598/barekengvol17iss2pp0969-0978.

[5] S. A. A. Kharis, G. F. Hertono, S. R. Irawan, E. Wahyuningrum, and Yumiati, "Students' success prediction based on the Fuzzy K-Nearest Neighbor method in Universitas Terbuka," *Education Technology in the New Normal: Now and Beyond*, pp. 212–218, Jun. 2023, doi: 10.1201/9781003353423-22.

[6] S. A. A. Kharis, A. Zili, E. Zubir, and F. Ihza Fajar, "Prediksi Kelulusan Siswa pada Mata Pelajaran Matematika menggunakan Educational Data Mining," *Jurnal Riset Pembelajaran Matematika Sekolah*, vol. 7, 2023, [Online]. Available: https://archive.ics.uci.edu/

[7] A. Haqqi *et al.*, "Peramalan Harga Saham Dengan Model Hybrid Arima-Garch dan Metode Walk Forward," *Jurnal Statistika dan Aplikasinya*, vol. 6, no. 2, 2022.

[8] D. Molodtsov, "Soft Set Theory First Results," 1999.

[9] S. M. Mostafa, F. F. Kareem, and H. A. Jad, "Brief Review of Soft Sets and Its Application in Coding Theory," pp. 95–106, 2020, [Online]. Available: http://www.newtheory.org

[10] Z. Xiao *et al.*, "A new evaluation method based on D-S generalized fuzzy soft sets and its application in medical diagnosis problem," *Appl Math Model*, vol. 36, no. 10, pp. 4592–4604, Oct. 2012, doi: 10.1016/j.apm.2011.11.049.

[11] M. Saqlain *et al.*, "A New Approach of Neutrosophic Soft Set with Generalized Fuzzy TOPSIS in Application of Smart Phone Selection," 2020.

[12] M. A. Qamar and N. Hassan, “An approach toward a Q-neutrosophic soft set and its application in decision making,” *Symmetry (Basel)*, vol. 11, no. 2, Feb. 2019, doi: 10.3390/sym11020139.

[13] F. Hao, D. S. Park, and Z. Pei, “When social computing meets soft computing: opportunities and insights,” *Human-centric Computing and Information Sciences*, vol. 8, no. 1, Dec. 2018, doi: 10.1186/s13673-018-0131-z.

[14] F. Fatimah, D. Rosadi, R. Fajriya Hakim, and J. R. Carlos Alcantud, “Probabilistic soft sets and dual probabilistic soft sets in decision making,” *J Phys Conf Ser*, 2018.

[15] S. M. Abbas, K. A. Alam, and S. Shamshirband, “A soft-rough set based approach for handling contextual sparsity in context-aware video recommender systems,” *Mathematics*, vol. 7, no. 8, Aug. 2019, doi: 10.3390/math7080740.

[16] K. Haruna *et al.*, “A Soft Set Approach for Handling Conflict Situation on Movie Selection,” *IEEE Access*, vol. 7, pp. 116179–116194, 2019, doi: 10.1109/ACCESS.2019.2892778.

[17] K. Haruna *et al.*, “A Soft Set Approach for Handling Conflict Situation on Movie Selection,” *IEEE Access*, vol. 7, pp. 116179–116194, 2019, doi: 10.1109/ACCESS.2019.2892778.

[18] T. Herawan and M. M. Deris, “A soft set approach for association rules mining,” *Knowl Based Syst*, vol. 24, no. 1, pp. 186–195, Feb. 2011, doi: 10.1016/j.knosys.2010.08.005.

[19] F. Feng, Q. Wang, R. R. Yager, J. C. José, and L. Zhang, “Maximal association analysis using logical formulas over soft sets,” *Expert Syst Appl*, vol. 159, p. 113557, Nov. 2020, doi: 10.1016/J.ESWA.2020.113557.

[20] F. Fatimah, D. Rosadi, R. B. F. Hakim, and J. C. R. Alcantud, “N-soft sets and their decision making algorithms,” *Soft comput*, vol. 22, no. 12, pp. 3829–3842, Jun. 2018, doi: 10.1007/S00500-017-2838-6/METRICS.

[21] F. Fatimah and J. Alcantud, “The Multi-Fuzzy N-Soft Set and its Applications to Decision-Making,” *Neural Comput Appl*, vol. 33, pp. 11437–11446, 2021.

[22] M. Akram, A. Adeel, and J. C. R. Alcantud, “Hesitant fuzzy N-soft sets: A new model with applications in decision-making,” *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 6, pp. 6113–6127, Jan. 2019, doi: 10.3233/JIFS-181972.

[23] F. Fatimah, D. Rosadi, R. B. F. Hakim, and J. C. R. Alcantud, “N-soft sets and their decision making algorithms,” *Soft comput*, vol. 22, no. 12, pp. 3829–3842, Jun. 2018, doi: 10.1007/S00500-017-2838-6/METRICS.

[24] C. Győrödi, “A Comparative Study of Distributed Algorithms in Mining Association Rules,” 2023.

[25] Q. Zhao, “Association Rule Mining: A Survey,” Singapore, 2003.

[26] W. Lin, S. A. Alvarez, and C. Ruiz, “Efficient Adaptive-Support Association Rule Mining for Recommender Systems,” 2002.

[27] T. Herawan and M. M. Deris, “A soft set approach for association rules mining,” *Knowl Based Syst*, vol. 24, no. 1, pp. 186–195, Feb. 2011, doi: 10.1016/j.knosys.2010.08.005.

[28] D. Molodtsov, “Soft Set Theory First Results,” *Computers and Mathematics with Applications*, vol. 37, pp. 19–31, 1999.

[29] F. Feng, Q. Wang, R. R. Yager, J. C. José, and L. Zhang, “Maximal association analysis using logical formulas over soft sets,” *Expert Syst Appl*, vol. 159, Nov. 2020, doi: 10.1016/j.eswa.2020.113557.