

Article

Neutrosophic Association Rule Mining Algorithm for Big Data Analysis

Mohamed Abdel-Basset ^{1,*} , Mai Mohamed ¹, Florentin Smarandache ^{2,*}  and Victor Chang ³

¹ Department of Operations Research, Faculty of Computers and Informatics, Zagazig University, Sharqiyah 44519, Egypt; analyst_mohamed@yahoo.com

² Math & Science Department, University of New Mexico, Gallup, NM 87301, USA

³ International Business School Suzhou, Xi'an Jiaotong-Liverpool University, Wuzhong, Suzhou 215123, China; ic.victor.chang@gmail.com

* Correspondence: analyst_mohamed@zu.edu.eg (M.A.-B.); smarand@unm.edu (F.S.)

Received: 5 March 2018; Accepted: 9 April 2018; Published: 11 April 2018



Abstract: Big Data is a large-sized and complex dataset, which cannot be managed using traditional data processing tools. Mining process of big data is the ability to extract valuable information from these large datasets. Association rule mining is a type of data mining process, which is indented to determine interesting associations between items and to establish a set of association rules whose support is greater than a specific threshold. The classical association rules can only be extracted from binary data where an item exists in a transaction, but it fails to deal effectively with quantitative attributes, through decreasing the quality of generated association rules due to sharp boundary problems. In order to overcome the drawbacks of classical association rule mining, we propose in this research a new neutrosophic association rule algorithm. The algorithm uses a new approach for generating association rules by dealing with membership, indeterminacy, and non-membership functions of items, conducting to an efficient decision-making system by considering all vague association rules. To prove the validity of the method, we compare the fuzzy mining and the neutrosophic mining. The results show that the proposed approach increases the number of generated association rules.

Keywords: neutrosophic association rule; data mining; neutrosophic sets; big data

1. Introduction

The term ‘Big Data’ originated from the massive amount of data produced every day. Each day, Google receives cca. 1 billion queries, Facebook registers more than 800 million updates, and YouTube counts up to 4 billion views, and the produced data grows with 40% every year. Other sources of data are mobile devices and big companies. The produced data may be structured, semi-structured, or unstructured. Most of the big data types are unstructured; only 20% of data consists in structured data. There are four dimensions of big data:

- (1) Volume: big data is measured by petabytes and zettabytes.
- (2) Velocity: the accelerating speed of data flow.
- (3) Variety: the various sources and types of data requiring analysis and management.
- (4) Veracity: noise, abnormality, and biases of generated knowledge.

Consequently, Gartner [1] outlines that big data’s large volume requires cost-effective, innovative forms for processing information, to enhance insights and decision-making processes.

Prominent domains among applications of big data are [2,3]:

- (1) Business domain.
- (2) Technology domain.
- (3) Health domain.
- (4) Smart cities designing.

These various applications help people to obtain better services, experiences, or be healthier, by detecting illness symptoms much earlier than before [2]. Some significant challenges of managing and analyzing big data are [4,5]:

- (1) Analytics Architecture: The optimal architecture for dealing with historic and real-time data at the same time is not obvious yet.
- (2) Statistical significance: Fulfill statistical results, which should not be random.
- (3) Distributed mining: Various data mining methods are not fiddling to paralyze.
- (4) Time evolving data: Data should be improved over time according to the field of interest.
- (5) Compression: To deal with big data, the amount of space that is needed to store is highly relevant.
- (6) Visualization: The main mission of big data analysis is the visualization of results.
- (7) Hidden big data: Large amounts of beneficial data are lost since modern data is unstructured data.

Due to the increasing volume of data at a matchless rate and of various forms, we need to manage and analyze uncertainty of various types of data. Big data analytics is a significant function of big data, which discovers unobserved patterns and relationships among various items and people interest on a specific item from the huge data set. Various methods are applied to obtain valid, unknown, and useful models from large data. Association rule mining stands among big data analytics functionalities. The concept of association rule (AR) mining already returns to H'ajek et al. [6]. Each association rule in database is composed from two different sets of items, which are called antecedent and consequent. A simple example of association rule mining is "if the client buys a fruit, he/she is 80% likely to purchase milk also". The previous association rule can help in making a marketing strategy of a grocery store. Then, we can say that association rule-mining finds all of the frequent items in database with the least complexities. From all of the available rules, in order to determine the rules of interest, a set of constraints must be determined. These constraints are support, confidence, lift, and conviction. Support indicates the number of occurrences of an item in all transactions, while the confidence constraint indicates the truth of the existing rule in transactions. The factor "lift" explains the dependency relationship between the antecedent and consequent. On the other hand, the conviction of a rule indicates the frequency ratio of an occurring antecedent without a consequent occurrence. Association rules mining could be limited to the problem of finding large itemsets, where a large itemset is a collection of items existing in a database transactions equal to or greater than the support threshold [7–20]. In [8], the author provides a survey of the itemset methods for discovering association rules. The association rules are positive and negative rules. The positive association rules take the form $X \rightarrow Y$, $X \subseteq I$, $Y \subseteq I$ and $X \cap Y = \varphi$, where X, Y are antecedent and consequent and I is a set of items in database. Each positive association rule may lead to three negative association rules, $\neg Y$, $X \rightarrow Y$, and $X \rightarrow Y$. Generating association rules in [9] consists of two problems. The first problem is to find frequent itemsets whose support satisfies a predefined minimum value. Then, the concern is to derive all of the rules exceeding a minimum confidence, based on each frequent itemset. Since the solution of the second problem is straightforward, most of the proposed work goes in for solving the first problem. An a priori algorithm has been proposed in [19], which was the basis for many of the forthcoming algorithms. A two-pass algorithm is presented in [11]. It consumes only two database scan passes, while a priori is a multi-pass algorithm and needs up to $c+1$ database scans, where c is the number of items (attributes). Association rules mining is applicable in numerous database communities. It has large applications in the retail industry to improve market basket analysis [7]. Streaming-Rules is an algorithm developed by [9] to report an association between pairs of elements in streams for predictive caching and detecting the previously

undetectable hit inflation attacks in advertising networks. Running mining algorithms on numerical attributes may result in a large set of candidates. Each candidate has small support and many rules have been generated with useless information, e.g., the age attribute, salary attribute, and students' grades. Many partitioning algorithms have been developed to solve the numerical attributes problem. The proposed algorithms faced two problems. The first problem was the partitioning of attribute domain into meaningful partitions. The second problem was the loss of many useful rules due to the sharp boundary problem. Consequently, some rules may fail to achieve the minimum support threshold because of the separating of its domain into two partitions.

Fuzzy sets have been introduced to solve these two problems. Using fuzzy sets make the resulted association rules more meaningful. Many mining algorithms have been introduced to solve the quantitative attributes problem using fuzzy sets proposed algorithms in [13–27] that can be separated into two types related to the kind of minimum support threshold, fuzzy mining based on single-minimum support threshold, and fuzzy mining based on multi-minimum support threshold [21]. Neutrosophic theory was introduced in [28] to generalize fuzzy theory. In [29–32], the neutrosophic theory has been proposed to solve several applications and it has been used to generate a solution based on neutrosophic sets. Single-valued neutrosophic set was introduced in [33] to transfer the neutrosophic theory from the philosophic field into the mathematical theory, and to become applicable in engineering applications. In [33], a differentiation has been proposed between intuitionistic fuzzy sets and neutrosophic sets based on the independence of membership functions (truth-membership function, falsity-membership function, and indeterminacy-membership function). In neutrosophic sets, indeterminacy is explicitly independent, and truth-membership function and falsity-membership function are independent as well. In this paper, we introduce an approach that is based on neutrosophic sets for mining association rules, instead of fuzzy sets. Also, a comparison resulted association rules in both of the scenarios has been presented. In [34], an attempt to express how neutrosophic sets theory could be used in data mining has been proposed. They define SVNSF (single-valued neutrosophic score function) to aggregate attribute values. In [35], an algorithm has been introduced to mining vague association rules. Items properties have been added to enhance the quality of mining association rules. In addition, almost sold items (items has been selected by the customer, but not checked out) were added to enhance the generated association rules. AH-pair Database consisting of a traditional database and the hesitation information of items was generated. The hesitation information was collected, depending online shopping stores, which make it easier to collect that type of information, which does not exist in traditional stores. In this paper, we are the first to convert numerical attributes (items) into neutrosophic sets. While vague association rules add new items from the hesitating information, our framework adds new items by converting the numerical attributes into linguistic terms. Therefore, the vague association rule mining can be run on the converted database, which contains new linguistic terms.

Research Contribution

Detecting hidden and affinity patterns from various, complex, and big data represents a significant role in various domain areas, such as marketing, business, medical analysis, etc. These patterns are beneficial for strategic decision-making. Association rules mining plays an important role as well in detecting the relationships between patterns for determining frequent itemsets, since classical association rules cannot use all types of data for the mining process. Binary data can only be used to form classical rules, where items either exist in database or not. However, when classical association rules deal with quantitative database, no discovered rules will appear, and this is the reason for innovating quantitative association rules. The quantitative method also leads to the sharp boundary problem, where the item is below or above the estimation values. The fuzzy association rules are introduced to overcome the classical association rules drawbacks. The item in fuzzy association rules has a membership function and a fuzzy set. The fuzzy association rules can deal with vague rules, but not in the best manner, since it cannot consider the indeterminacy of rules. In order to overcome

drawbacks of previous association rules, a new neutrosophic association rule algorithm has been introduced in this research. Our proposed algorithm deals effectively and efficiently with vague rules by considering not only the membership function of items, but also the indeterminacy and the falsity functions. Therefore, the proposed algorithm discovers all of the possible association rules and minimizes the losing processes of rules, which leads to building efficient and reliable decision-making system. By comparing our proposed algorithm with fuzzy approaches, we note that the number of association rules is increased, and negative rules are also discovered. The separation of negative association rules from positive ones is not a simple process, and it helps in various fields. As an example, in the medical domain, both positive and negative association rules help not only in the diagnosis of disease, but also in detecting prevention manners.

The rest of this research is organized as follows. The basic concepts and definitions of association rules mining are presented in Section 2. A quick overview of fuzzy association rules is described in Section 3. The neutrosophic association rules and the proposed model are presented in Section 4. A case study of Telecom Egypt Company is presented in Section 5. The experimental results and comparisons between fuzzy and proposed association rules are discussed in Section 6. The conclusions are drawn in Section 7.

2. Association Rules Mining

In this section, we formulate the $|D|$ transactions from the mining association rules for a database D . We used the following notations:

- (i) $I = \{i_1, i_2, \dots, i_m\}$ represents all the possible data sets, called items.
- (ii) Transaction set T is the set of domain data resulting from transactional processing such as $T \subseteq I$.
- (iii) For a given itemset $X \subseteq I$ and a given transaction T , we say that T contains X if and only if $X \subseteq T$.
- (iv) σ_X : the support frequency of X , which is defined as the number of transactions out of D that contain X .
- (v) s : the support threshold.

X is considered a large itemset, if $\sigma_X \geq |D| \times s$. Further, an association rule is an implication of the form $X \Rightarrow Y$, where $X \subseteq I$, $Y \subseteq I$ and $X \cap Y = \varphi$.

An association rule $X \Rightarrow Y$ is addressed in D with confidence c if at least c transactions out of D contain both X and Y . The rule $X \Rightarrow Y$ is considered as a large itemset having a minimum support s if: $\sigma_{X \cup Y} \geq |D| \times s$.

For a specific confidence and specific support thresholds, the problem of mining association rules is to find out all of the association rules having confidence and support that is larger than the corresponding thresholds. This problem can simply be expressed as finding all of the large itemsets, where a large itemset L is:

$$L = \{X | X \subseteq I \wedge \sigma_X \geq |D| \times s\}.$$

3. Fuzzy Association Rules

Mining of association rules is considered as the main task in data mining. An association rule expresses an interesting relationship between different attributes. Fuzzy association rules can deal with both quantitative and categorical data and are described in linguistic terms, which are understandable terms [26].

Let $T = \{t_1, \dots, t_n\}$ be a database transactions. Each transaction consists of a number of attributes (items). Let $I = \{i_1, \dots, i_m\}$ be a set of categorical or quantitative attributes. For each attribute i_k , ($k = 1, \dots, m$), we consider $\{n_1, \dots, n_k\}$ associated fuzzy sets. Typically, a domain expert determines the membership function for each attribute.

The tuple $\langle X, A \rangle$ is called the fuzzy itemset, where $X \subseteq I$ (set of attributes) and A is a set of fuzzy sets that is associated with attributes from X .

Following is an example of fuzzy association rule:

IF salary is high and age is old THEN insurance is high

Before the mining process starts, we need to deal with numerical attributes and prepare them for the mining process. The main idea is to determine the linguistic terms for the numerical attribute and define the range for every linguistic term. For example, the temperature attribute is determined by the linguistic terms {very cold, cold, cool, warm, hot}. Figure 1 illustrates the membership function of the temperature attribute.

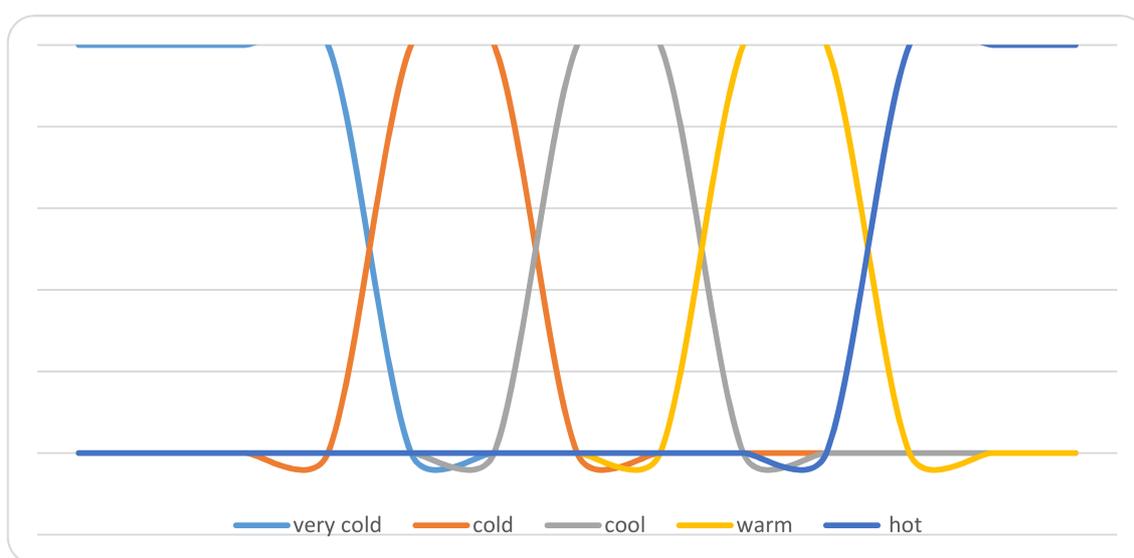


Figure 1. Linguistic terms of the temperature attribute.

The membership function has been calculated for the following database transactions illustrated in Table 1.

Table 1. Membership function for Database Transactions.

Transaction	Temp.	Membership Degree
T1	18	1 cool
T2	13	0.6 cool, 0.4 cold
T3	12	0.4 cool, 0.6 cold
T4	33	0.6 warm, 0.4 hot
T5	21	0.2 warm, 0.8 cool
T6	25	1 warm

We add the linguistic terms {very cold, cold, cool, warm, hot} to the candidate set and calculate the support for those itemsets. After determining the linguistic terms for each numerical attribute, the fuzzy candidate set have been generated.

Table 2 contains the support for each itemset individual one-itemsets. The count for every linguistic term has been calculated by summing its membership degree over the transactions. Table 3 shows the support for two-itemsets. The count for the fuzzy sets is the summation of degrees that resulted from the membership function of that itemset. The count for two-itemset has been calculated by summing the minimum membership degree of the 2 items. For example, {cold, cool} has count 0.8, which resulted from transactions T2 and T3. For transaction T2, membership degree of cool is 0.6 and membership degree for cold is 0.4, so the count for set {cold, cool} in T2 is 0.4. Also, T3 has the same count for {cold, cool}. So, the count of set {cold, cool} over all transactions is 0.8.

Table 2. 1-itemset support.

1-itemset	Count	Support
Very cold	0	0
Cold	1	0.17
Cool	2.8	0.47
Warm	1.6	0.27
Hot	0.6	0.1

Table 3. 2-itemset support.

2-itemset	Count	Support
{Cold, cool}	0.8	0.13
{Warm, hot}	0.4	0.07
{warm, cool}	0.2	0.03

In subsequent discussions, we denote an itemset that contains k items as k -itemset. The set of all k -itemsets in L is referred as L_k .

4. Neutrosophic Association Rules

In this section, we overview some basic concepts of the NSs and SVNNSs over the universal set X , and the proposed model of discovering neutrosophic association rules.

4.1. Neutrosophic Set Definitions and Operations

Definition 1 ([33]). Let X be a space of points and $x \in X$. A neutrosophic set (NS) A in X is definite by a truth-membership function $T_A(x)$, an indeterminacy-membership function $I_A(x)$ and a falsity-membership function $F_A(x)$. $T_A(x)$, $I_A(x)$ and $F_A(x)$ are real standard or real nonstandard subsets of $]^-0, 1^+[$. That is $T_A(x): X \rightarrow]^-0, 1^+[$, $I_A(x): X \rightarrow]^-0, 1^+[$ and $F_A(x): X \rightarrow]^-0, 1^+[$. There is no restriction on the sum of $T_A(x)$, $I_A(x)$ and $F_A(x)$, so $0^- \leq \sup T_A(x) + \sup I_A(x) + \sup F_A(x) \leq 3^+$.

Neutrosophic is built on a philosophical concept, which makes it difficult to process during engineering applications or to use it to real applications. To overcome that, Wang et al. [31], defined the SVNNS, which is a particular case of NS.

Definition 2. Let X be a universe of discourse. A single valued neutrosophic set (SVNS) A over X is an object taking the form $A = \{ \langle x, T_A(x), I_A(x), F_A(x) \rangle : x \in X \}$, where $T_A(x): X \rightarrow [0, 1]$, $I_A(x): X \rightarrow [0, 1]$ and $F_A(x): X \rightarrow [0, 1]$ with $0 \leq T_A(x) + I_A(x) + F_A(x) \leq 3$ for all $x \in X$. The intervals $T_A(x)$, $I_A(x)$ and $F_A(x)$ represent the truth-membership degree, the indeterminacy-membership degree and the falsity membership degree of x to A , respectively. For convenience, a SVN number is represented by $A = (a, b, c)$, where $a, b, c \in [0, 1]$ and $a + b + c \leq 3$.

Definition 3 (Intersection) ([31]). For two SVNNSs $A = \langle T_A(x), I_A(x), F_A(x) \rangle$ and $B = \langle T_B(x), I_B(x), F_B(x) \rangle$, the intersection of these SVNNSs is again an SVNNS which is defined as $C = A \cap B$ whose truth, indeterminacy and falsity membership functions are defined as $T_C(x) = \min(T_A(x), T_B(x))$, $I_C(x) = \max(I_A(x), I_B(x))$ and $F_C(x) = \max(F_A(x), F_B(x))$.

Definition 4 (Union) ([31]). For two SVNNSs $A = \langle T_A(x), I_A(x), F_A(x) \rangle$ and $B = \langle T_B(x), I_B(x), F_B(x) \rangle$, the union of these SVNNSs is again an SVNNS which is defined as $C = A \cup B$ whose truth, indeterminacy and falsity membership functions are defined as $T_C(x) = \max(T_A(x), T_B(x))$, $I_C(x) = \min(I_A(x), I_B(x))$ and $F_C(x) = \min(F_A(x), F_B(x))$.

Definition 5 (Containment) ([31]). A single valued neutrosophic set A contained in the other SVN B , denoted by $A \subseteq B$ if and only if $T_A(x) \leq T_B(x)$, $I_A(x) \leq I_B(x)$ and $F_A(x) \geq F_B(x)$ for all x in X .

Next, we propose a method for generating the association rule under the SVN environment.

4.2. Proposed Model for Association Rule

In this paper, we introduce a model to generate association rules of form:

$$X \rightarrow Y \text{ where } X \cap Y = \varphi \text{ and } X, Y \text{ are neutrosophic sets.}$$

Our aim is to find the frequent itemsets and their corresponding support. Generating an association rule from its frequent itemsets, which are dependent on the confidence threshold, are also discussed here. This has been done by adding the neutrosophic set into I , where I is all of the possible data sets, which are referred as items. So $I = N \cup M$ where N is neutrosophic set and M is classical set of items. The general form of an association rule is an implication of the form $X \rightarrow Y$, where $X \subseteq I$, $Y \subseteq I$, $X \cap Y = \varphi$.

Therefore, an association rule $X \rightarrow Y$ is addressed in Database D with confidence ‘ c ’ if at least c transactions out of D contains both X and Y . On the other hand, the rule $X \rightarrow Y$ is considered a large item set having a minimum support s if $\sigma_{X \cup Y} \geq |D| \times s$. Furthermore, the process of converting the quantitative values into the neutrosophic sets is proposed, as shown in Figure 2.

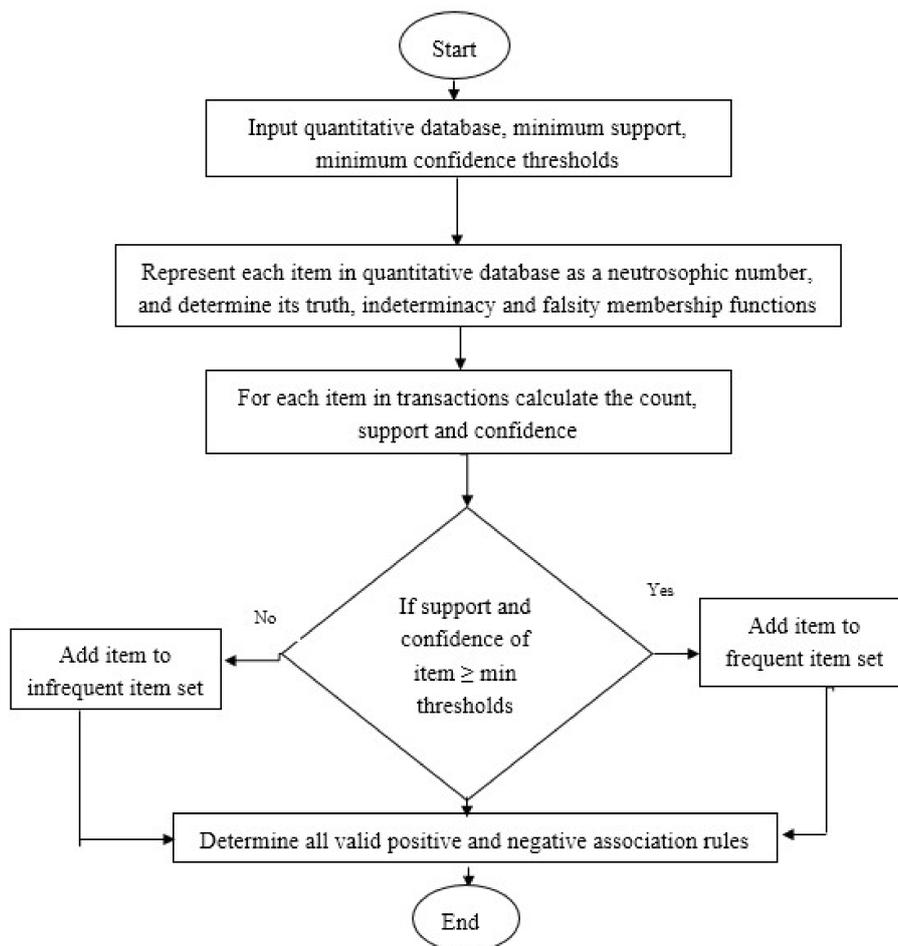


Figure 2. The proposed model.

The proposed model for the construction of the neutrosophic numbers is summarized in the following steps:

- Step 1** Set linguistic terms of the variable, which will be used for quantitative attribute.
- Step 2** Define the truth, indeterminacy, and the falsity membership functions for each constructed linguistic term.
- Step 3** For each transaction t in T , compute the truth-membership, indeterminacy-membership and falsity-membership degrees.
- Step 4** Extend each linguistic term l in set of linguistic terms L into T_L , I_L , and F_L to denote truth-membership, indeterminacy-membership, and falsity-membership functions, respectively.
- Step 5** For each k -item set where $k = \{1, 2, \dots, n\}$, and n number of iterations.
 - calculate count of each linguistic term by summing degrees of membership for each transaction as $Count(A) = \sum_{i=1}^{i=t} \mu_A(x)$ where μ_A is T_A, I_A or F_A .
 - calculate support for each linguistic term $s = \frac{Count(A)}{No. \text{ of transactions}}$.

Step 6 The above procedure has been repeated for every quantitative attribute in the database.

In order to show the working procedure of the approach, we consider the temperature as an attribute and the terms “very cold”, “cold”, “cool”, “warm”, and “hot” as their linguistic terms to represent the temperature of an object. Then, following the steps of the proposed approach, construct their membership function as below:

Step 1 The attribute temperature’ has set the linguistic terms “very cold”, “cold”, “cool”, “warm”, and “hot”, and their ranges are defined in Table 4.

Table 4. Linguistic terms ranges.

Linguistic Term	Core Range	Left Boundary Range	Right Boundary Range
Very Cold	$-\infty-0$	N/A	0–5
Cold	5–10	0–5	10–15
Cool	15–20	10–15	20–25
Warm	25–30	20–25	30–35
Hot	35– ∞	30–35	N/A

Step 2 Based on these linguistic term ranges, the truth-membership functions of each linguistic variable are defined, as follows:

$$T_{very-cold}(x) = \begin{cases} 1 & ; \text{for } x \leq 0 \\ (5 - x)/5 & ; \text{for } 0 < x < 5 \\ 0 & ; \text{for } x \geq 5 \end{cases}$$

$$T_{cold}(x) = \begin{cases} 1 & ; \text{for } 5 \leq x \leq 10 \\ (15 - x)/5 & ; \text{for } 10 < x < 15 \\ x/5 & ; \text{for } 0 < x < 5 \\ 0 & ; \text{for } x \geq 15 \text{ or } x \leq 0 \end{cases}$$

$$T_{cool}(x) = \begin{cases} 1 & ; \text{for } 15 \leq x \leq 20 \\ (25 - x)/5 & ; \text{for } 20 < x < 25 \\ (x - 10)/5 & ; \text{for } 10 < x < 15 \\ 0 & ; \text{otherwise} \end{cases}$$

$$T_{warm}(x) = \begin{cases} 1 & ; \text{for } 25 \leq x \leq 30 \\ (35 - x)/5 & ; \text{for } 30 < x < 35 \\ (x - 20)/5 & ; \text{for } 20 < x < 25 \\ 0 & ; \text{otherwise} \end{cases}$$

$$T_{hot}(x) = \begin{cases} 1 & ; \text{for } x \geq 35 \\ (x - 30)/5 & ; \text{for } 30 < x < 35 \\ 0 & ; \text{otherwise} \end{cases}$$

The falsity-membership functions of each linguistic variable are defined as follows:

$$F_{very-cold}(x) = \begin{cases} 0 & ; \text{for } x \leq 0 \\ x/5 & ; \text{for } 0 < x < 5 \\ 1 & ; \text{for } x \geq 5 \end{cases}$$

$$F_{cold}(x) = \begin{cases} 0 & ; \text{for } 5 \leq x \leq 10 \\ (x - 10)/5 & ; \text{for } 10 < x < 15 \\ (5 - x)/5 & ; \text{for } 0 < x < 5 \\ 1 & ; \text{for } x \geq 15 \text{ or } x \leq 0 \end{cases}$$

$$F_{cool}(x) = \begin{cases} 0 & ; \text{for } 15 \leq x \leq 20 \\ (x - 20)/5 & ; \text{for } 20 < x < 25 \\ (15 - x)/5 & ; \text{for } 10 < x < 15 \\ 1 & ; \text{otherwise} \end{cases}$$

$$F_{warm}(x) = \begin{cases} 0 & ; \text{for } 25 \leq x \leq 30 \\ (x - 30)/5 & ; \text{for } 30 < x < 35 \\ (25 - x)/5 & ; \text{for } 20 < x < 25 \\ 1 & ; \text{otherwise} \end{cases}$$

$$F_{hot}(x) = \begin{cases} 0 & ; \text{for } x \geq 35 \\ (35 - x)/5 & ; \text{for } 30 < x < 35 \\ 1 & ; \text{otherwise} \end{cases}$$

The indeterminacy membership functions of each linguistic variables are defined as follows:

$$I_{very-cold}(x) = \begin{cases} 0 & ; \text{for } x \leq -2.5 \\ (x + 2.5)/5 & ; \text{for } -2.5 \leq x \leq 2.5 \\ (7.5 - x)/5 & ; \text{for } 2.5 \leq x \leq 7.5 \\ 0 & ; \text{for } x \geq 7.5 \end{cases}$$

$$I_{cold}(x) = \begin{cases} (x + 2.5)/5 & ; \text{for } 2.5 \leq x \leq 2.5 \\ (7.5 - x)/5 & ; \text{for } 2.5 \leq x \leq 7.5 \\ (x - 7.5)/5 & ; \text{for } 7.5 \leq x \leq 12.5 \\ (17.5 - x)/5 & ; \text{for } 12.5 \leq x \leq 17.5 \\ 0 & ; \text{otherwise} \end{cases}$$

$$I_{cool}(x) = \begin{cases} (x - 7.5)/5 & ; \text{for } 7.5 \leq x \leq 12.5 \\ (17.5 - x)/5 & ; \text{for } 12.5 \leq x \leq 17.5 \\ (x - 17.5)/5 & ; \text{for } 17.5 \leq x \leq 22.5 \\ (27.5 - x)/5 & ; \text{for } 22.5 \leq x \leq 27.5 \\ 0 & ; \text{otherwise} \end{cases}$$

$$I_{warm}(x) = \begin{cases} (x - 17.5)/5 & ; \text{for } 17.5 \leq x \leq 22.5 \\ (27.5 - x)/5 & ; \text{for } 22.5 \leq x \leq 27.5 \\ (x - 27.5)/5 & ; \text{for } 27.5 \leq x \leq 32.5 \\ (37.5 - x)/5 & ; \text{for } 32.5 \leq x \leq 37.5 \\ 0 & ; \text{otherwise} \end{cases}$$

$$I_{hot}(x) = \begin{cases} (x - 27.5)/5 & ; \text{for } 27.5 \leq x \leq 32.5 \\ (37.5 - x)/5 & ; \text{for } 32.5 \leq x \leq 37.5 \\ 0 & ; \text{otherwise} \end{cases}$$

The graphical membership degrees of these variables are summarized in Figure 3. The graphical falsity degrees of these variables are summarized in Figure 4. Also, the graphical indeterminacy degrees of these variables are summarized in Figure 5. On the other hand, for a particular linguistic term, ‘Cool’ in the temperature attribute, their neutrosophic membership functions are represented in Figure 6.



Figure 3. Truth-membership function of temperature attribute.

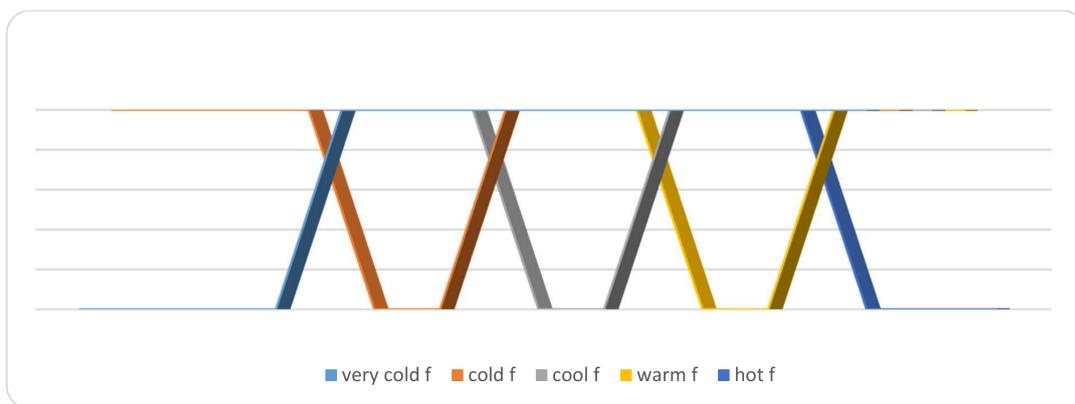


Figure 4. Falsity-membership function of temperature attribute.

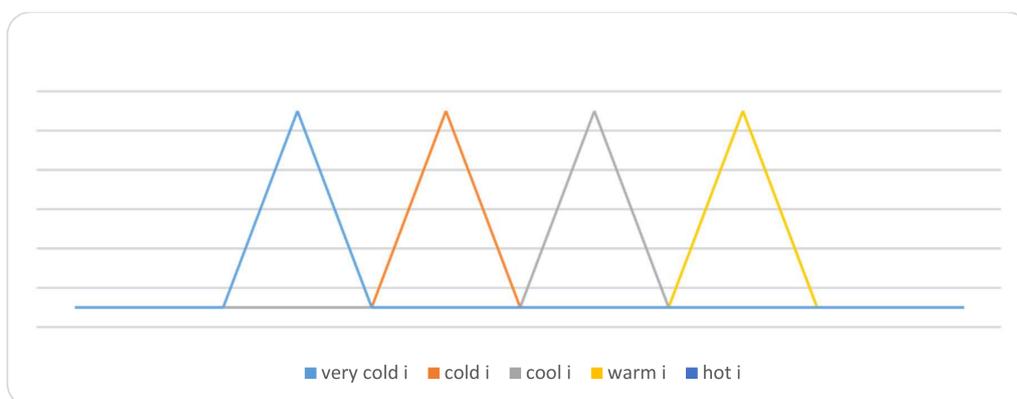


Figure 5. Indeterminacy-membership function of temperature attribute.

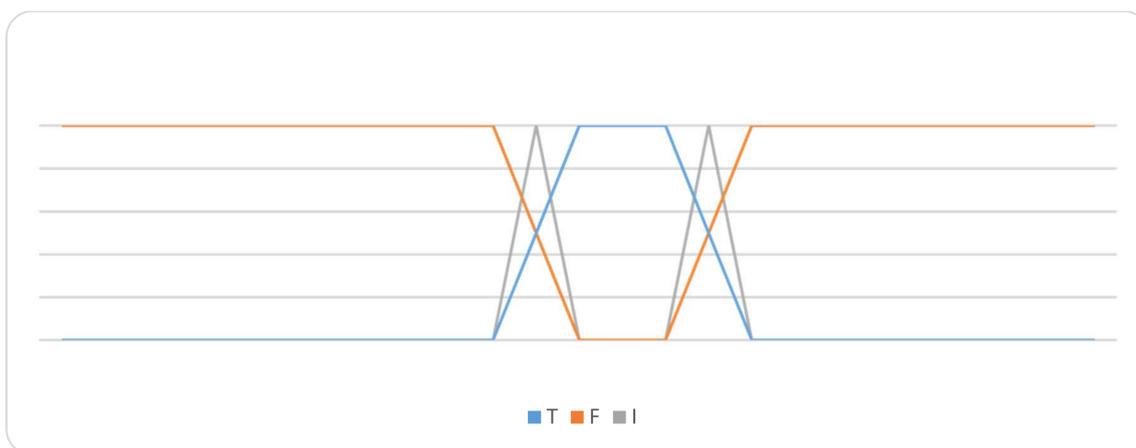


Figure 6. Cool (T, I, F) for temperature attribute.

Step 3 Based on the membership grades, different transaction has been set up by taking different sets of the temperatures. The membership grades in terms of the neutrosophic sets of these transactions are summarized in Table 5.

Table 5. Membership function for database Transactions.

Transaction	Temp.	Membership Degree
T1	18	Very-cold <0,0,1> cold <0,0,1> cool <1,0.1,0> warm <0,0.1,1> hot <0,0,1>
T2	13	Very cold <0,0,1> cold <0.4,0.9,0.6> cool <0.6,0.9,0.4> warm <0,0,1> hot <0,0,1>
T3	12	Very cold <0,0,1> cold <0.6,0.9,0.4> cool <0.4,0.9,0.6> warm <0,0,1> hot <0,0,1>
T4	33	Very cold <0,0,1> cold <0,0,1> cool <0,0,1> warm <0.4,0.9,0.6> hot <0.6,0.9,0.4>
T5	21	Very cold <0,0,1> cold <0,0,1> cool <0.8,0.7,0.2> warm <0.2,0.7,0.8> hot <0,0,1>
T6	25	Very cold <0,0,1> cold <0,0,1> cool <0,0,1> warm <1,0.5,0> hot <0,0,1>

- Step 4** Now, we count the set of linguistic terms {very cold, cold, cool, warm, hot} for every element in transactions. Since the truth, falsity, and indeterminacy-memberships are independent functions, the set of linguistic terms can be extended to $\{T_{very-cold}, T_{cold}, T_{cool}, T_{warm}, T_{hot}, F_{very-cold}, F_{cold}, F_{cool}, F_{warm}, F_{hot}, I_{very-cold}, I_{cold}, I_{cool}, I_{warm}, I_{hot}\}$ where F_{warm} means not worm and I_{warm} means not sure of warmness. This enhances dealing with negative association rules, which is handled as positive rules without extra calculations.
- Step 5** By using the membership degrees that are given in Table 5 for candidate 1-itemset, the count and support has been calculated, respectively. The corresponding results are summarized in Table 6.

Table 6. Support for candidate 1-itemset neutrosophic set.

1-itemset	Count	Support
$T_{verycold}$	0	0
T_{Cold}	1	0.17
T_{Cool}	2.8	0.47
T_{Warm}	1.6	0.27
T_{Hot}	0.6	0.1
$I_{verycold}$	0	0
I_{Cold}	1.8	0.3
I_{Cool}	2.6	0.43
I_{Warm}	2.2	0.37
I_{Hot}	0.9	0.15
$F_{verycold}$	6	1
F_{Cold}	5	0.83
F_{Cool}	3.2	0.53
F_{Warm}	4.4	0.73
F_{Hot}	5.4	0.9

Similarly, the two-itemset support is illustrated in Table 7 and the rest of itemset generation (k -itemset for $k = 3, 4 \dots 8$) are obtained similarly. The count for k -item set in database record is defined by minimum count of each one-itemset exists.

For example: $\{T_{Cold}, T_{Cool}\}$ count is 0.8

Because they exists in both T2 and T3.

In T2: $T_{Cold} = 0.4$ and $T_{Cool} = 0.6$ so, count for $\{T_{Cold}, T_{Cool}\}$ in T2 = 0.4

In T3: $T_{Cold} = 0.6$ and $T_{Cool} = 0.4$ so, count for $\{T_{Cold}, T_{Cool}\}$ in T2 = 0.4

Thus, count of $\{T_{Cold}, T_{Cool}\}$ in (Database) DB is 0.8.

Table 7. Support for candidate 2-itemset neutrosophic set.

2-itemset	Count	Support	2-itemset	Count	Support
$\{T_{Cold}, T_{Cool}\}$	0.8	0.13	$\{I_{Cold}, I_{Cool}\}$	1.8	0.30
$\{T_{Cold}, I_{Cold}\}$	1	0.17	$\{I_{Cold}, F_{verycold}\}$	1.8	0.30
$\{T_{Cold}, I_{Cool}\}$	1	0.17	$\{I_{Cold}, F_{Cold}\}$	1	0.17
$\{T_{Cold}, F_{verycold}\}$	1	0.17	$\{I_{Cold}, F_{Cool}\}$	1	0.17
$\{T_{Cold}, F_{Cold}\}$	0.8	0.13	$\{I_{Cold}, F_{Warm}\}$	1.8	0.30
$\{T_{Cold}, F_{Cool}\}$	1	0.17	$\{I_{Cold}, F_{Hot}\}$	1.8	0.30
$\{T_{Cold}, F_{Warm}\}$	1	0.17	$\{I_{Cool}, I_{Warm}\}$	0.8	0.13
$\{T_{Cold}, F_{Hot}\}$	1	0.17	$\{I_{Cool}, F_{verycold}\}$	2.6	0.43
$\{T_{Cool}, T_{Warm}\}$	0.2	0.03	$\{I_{Cool}, F_{Cold}\}$	1.8	0.30
$\{T_{Cool}, I_{Cold}\}$	1	0.17	$\{I_{Cool}, F_{Cool}\}$	1.2	0.20
$\{T_{Cool}, F_{Cool}\}$	1.8	0.30	$\{I_{Cool}, F_{Warm}\}$	2.6	0.43
$\{T_{Cool}, I_{Warm}\}$	0.8	0.13	$\{I_{Cool}, F_{Hot}\}$	2.6	0.43
$\{T_{Cool}, F_{verycold}\}$	2.8	0.47	$\{I_{Warm}, I_{Hot}\}$	0.9	0.15

Table 7. Cont.

{T _{Cool} , F _{Cold} }	2.8	0.47	{I _{Warm} , F _{verycold} }	2.2	0.37
{T _{Cool} , F _{Cool} }	1	0.17	{I _{Warm} , F _{Cold} }	2.2	0.37
{T _{Cool} , F _{Warm} }	2.8	0.47	{I _{Warm} , F _{Cool} }	1.6	0.27
{T _{Cool} , F _{Hot} }	2.8	0.47	{I _{Warm} , F _{Warm} }	1.4	0.23
{T _{Warm} , T _{Hot} }	0.4	0.07	{I _{Warm} , F _{Hot} }	1.7	0.28
{T _{Warm} , I _{Cool} }	0.2	0.03	{I _{Hot} , F _{verycold} }	0.9	0.15
{T _{Warm} , I _{Warm} }	1.1	0.18	{I _{Hot} , F _{Cold} }	0.9	0.15
{T _{Warm} , I _{Hot} }	0.4	0.07	{I _{Hot} , F _{Cool} }	0.9	0.15
{T _{Warm} , F _{verycold} }	1.6	0.27	{I _{Hot} , F _{Warm} }	0.6	0.10
{T _{Warm} , F _{Cold} }	1.6	0.27	{I _{Hot} , F _{Hot} }	0.4	0.07
{T _{Warm} , F _{Cool} }	1.6	0.27	{F _{verycold} , F _{Cold} }	5	0.83
{T _{Warm} , F _{Warm} }	0.6	0.10	{F _{verycold} , F _{Cool} }	3.2	0.53
{T _{Warm} , F _{Hot} }	1.6	0.27	{F _{verycold} , F _{Warm} }	4.4	0.73
{T _{Hot} , I _{Warm} }	0.6	0.10	{F _{verycold} , F _{Hot} }	5.4	0.90
{T _{Hot} , I _{Hot} }	0.6	0.10	{F _{Cold} , F _{Cool} }	3	0.50
{T _{Hot} , F _{verycold} }	0.6	0.10	{F _{Cold} , F _{Warm} }	3.4	0.57
{T _{Hot} , F _{Cold} }	0.6	0.10	{F _{Cold} , F _{Hot} }	4.4	0.73
{T _{Hot} , F _{Cool} }	0.6	0.10	{F _{Cool} , F _{Warm} }	1.8	0.30
{T _{Hot} , F _{Warm} }	0.6	0.10	{F _{Cool} , F _{Hot} }	2.6	0.43
{T _{Hot} , F _{Hot} }	0.4	0.07	{F _{Warm} , F _{Hot} }	4.2	0.70

5. Case Study

In this section, the case of Telecom Egypt Company stock records has been studied. Egyptian stock market has many companies. One of the major questions for stock market users is when to buy or to sell a specific stock. Egyptian stock market has three indicators, EGX30, EGX70, and EGX100. Each indicator gives a reflection of the stock market. Also, these indicators have an important impact on the stock market users, affecting their decisions of buying or selling stocks. We focus in our study on the relation between the stock and the three indicators. Also, we consider the month and quarter of the year to be another dimension in our study, while the sell/buy volume of the stock per day is considered to be the third dimension.

In this study, the historical data has been taken from the Egyptian stock market program (Mist) during the program September 2012 until September 2017. For every stock/indicator, Mist keeps a daily track of number of values (opening price, closing price, high price reached, low price reached, and volume). The collected data of Telecom Egypt Stock are summarized in Figure 7.



Symbolid	Ts_date	Open	Close	High	Low	Volume
S48031C016	012-09-13	13.98	14.07	14.29	13.85	490,997
S48031C016	012-09-16	14.08	14.09	14.20	14.02	347,258
S48031C016	012-09-17	14.11	14.60	15.01	14.11	2,050,226
S48031C016	012-09-18	14.52	14.72	15.00	14.52	500,290
S48031C016	012-09-19	14.90	14.68	14.95	14.41	362,893
S48031C016	012-09-20	14.77	14.84	15.00	14.50	913,683
S48031C016	012-09-23	14.56	15.09	15.37	14.56	283,008
S48031C016	012-09-24	14.80	14.73	15.00	14.68	896,914
S48031C016	012-09-25	14.80	14.39	14.80	14.00	2,604,812
S48031C016	012-09-26	14.25	14.53	14.71	14.25	648,102

Figure 7. Telecom Egypt stock records.

In this study, we use the open price and close price values to get price change rate, which are defined as follows:

$$\text{price change rate} = \frac{\text{close price} - \text{open price}}{\text{open price}} \times 100$$

and change the volume to be a percentage of total volume of the stock with the following relation:

$$\text{percentage of volume} = \frac{\text{volume}}{\text{total volume}} \times 100$$

The same was performed for the stock market indicators. Now, we take the attributes as “quarter”, “month”, “stock change rate”, “volume percentage”, and “indicators change rate”. Table 8 illustrates the segment of resulted data after preparation.

Table 8. Segment of data after preparation.

Ts_Date	Month	Quarter	Change	Volume	Change30	Change70	Change100
13 September 2012	September	3	0.64	0.03	−1.11	0.01	−0.43
16 September 2012	September	3	0.07	0.02	2.82	4.50	3.67
17 September 2012	September	3	3.47	0.12	1.27	0.76	0.81
18 September 2012	September	3	1.38	0.03	−0.08	−0.48	−0.43
19 September 2012	September	3	−1.48	0.02	0.35	−1.10	−0.64
20 September 2012	September	3	0.47	0.05	−1.41	−1.64	−1.55
23 September 2012	September	3	3.64	0.02	−0.21	1.00	0.41
24 September 2012	September	3	−0.47	0.05	0.27	−0.09	0.03
25 September 2012	September	3	−2.77	0.15	2.15	1.79	1.85
26 September 2012	September	3	1.96	0.04	0.22	0.96	0.57
27 September 2012	September	3	0.90	0.05	−1.38	−0.88	−0.92
30 September 2012	September	3	−0.14	0.00	−1.11	−0.79	−0.75
1 October 2012	October	4	−1.60	0.02	−2.95	−4.00	−3.51

Based on these linguistic terms, define the ranges under the SVN environment. For this, corresponding to the attribute in “change rate” and “volume”, the truth-membership functions by defining their linguistic terms as {“high up”, “high low”, “no change”, “low down”, “high down”} corresponding to attribute “change rate”, while for the attribute “volume”, the linguistic terms are (low, medium, and high) and their ranges are summarized in Figures 8 and 9, respectively. The falsity-membership function and indeterminacy-membership function have been calculated and applied as well for change rate attribute.

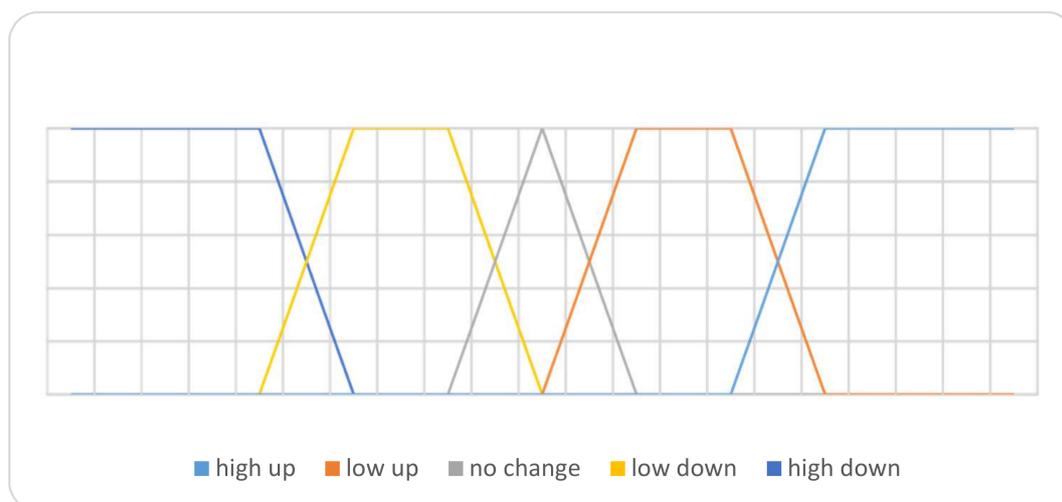


Figure 8. Change rate attribute truth-membership function.

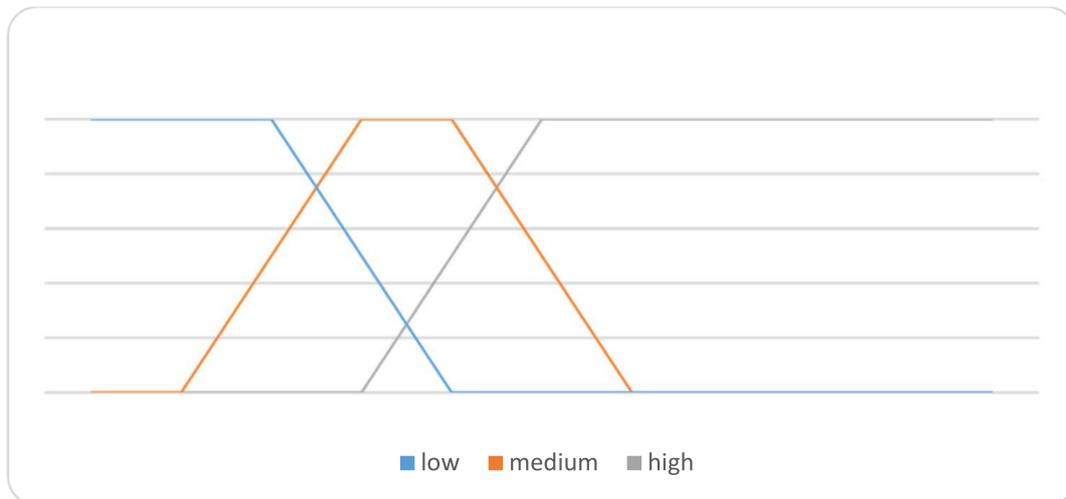


Figure 9. Volume attribute truth-membership function.

6. Experimental Results

We proceeded to a comparison between fuzzy mining and neutrosophic mining algorithms, and we found out that the number of generated association rules increased in neutrosophic mining.

A program has been developed to generate large itemsets for Telecom Egypt historical data. VB.net has been used in creating this program. The obtained data have been stored in an access database. The comparison depends on the number of generated association rules in a different min-support threshold. It should be noted that the performance cannot be part of the comparison because of the number of items (attributes) that are different in fuzzy vs. neutrosophic association rules mining. In fuzzy mining, the number of items was 14, while in neutrosophic mining it is 34. This happens because the number of attributes increased. Spreading each linguistic term into three (True, False, Indeterminacy) terms make the generated rules increase. The falsity-generated association rules can be considered a negative association rules. As pointed out in [36], the conviction of a rule $conv(X \rightarrow Y)$ is defined as the ratio of the expected frequency that X happened without Y falsity-association rules to be used to generate negative association rules if $T(x) + F(x) = 1$. In Table 9, the number of generated fuzzy rules in each k -itemset using different min-support threshold are reported, while the total generated fuzzy association rule is presented in Figure 10.

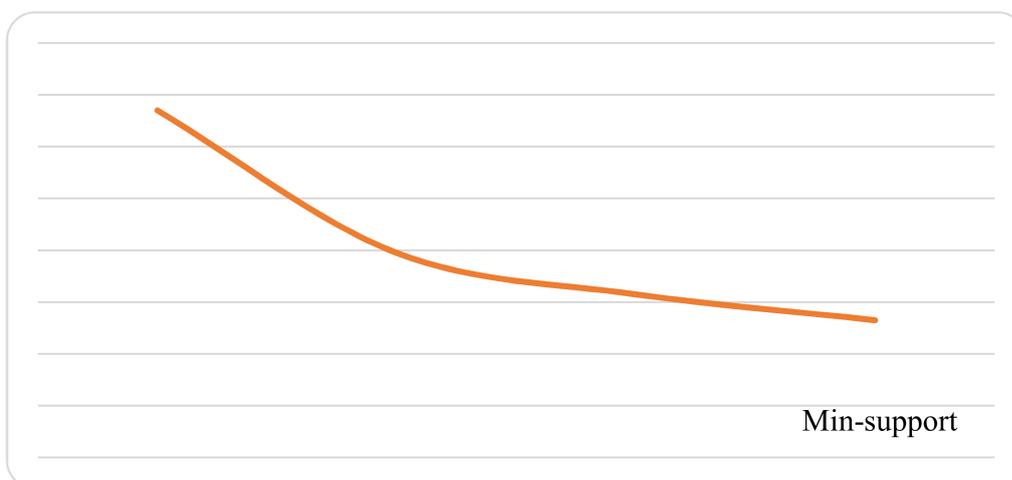


Figure 10. No. of fuzzy association rules with different min-support threshold.

Table 9. No. of resulted fuzzy rules with different min-support.

Min-Support	0.02	0.03	0.04	0.05
1-itemset	10	10	10	10
2-itemset	37	36	36	33
3-itemset	55	29	15	10
4-itemset	32	4	2	0

As compared to the fuzzy approach, by applying the same min-support threshold, we get a huge set of neutrosophic association rules. Table 10 illustrates the booming that happened to generated neutrosophic association rules. We stop generating itemsets at iteration 4 due to the noted expansion in the results shown in Figure 11, which shows the number of neutrosophic association rules.

Table 10. No. of neutrosophic rules with different min-support threshold.

Min-Support	0.02	0.03	0.04	0.05
1-itemset	26	26	26	26
2-itemset	313	311	309	300
3-itemset	2293	2164	2030	1907
4-itemset	11,233	9689	8523	7768

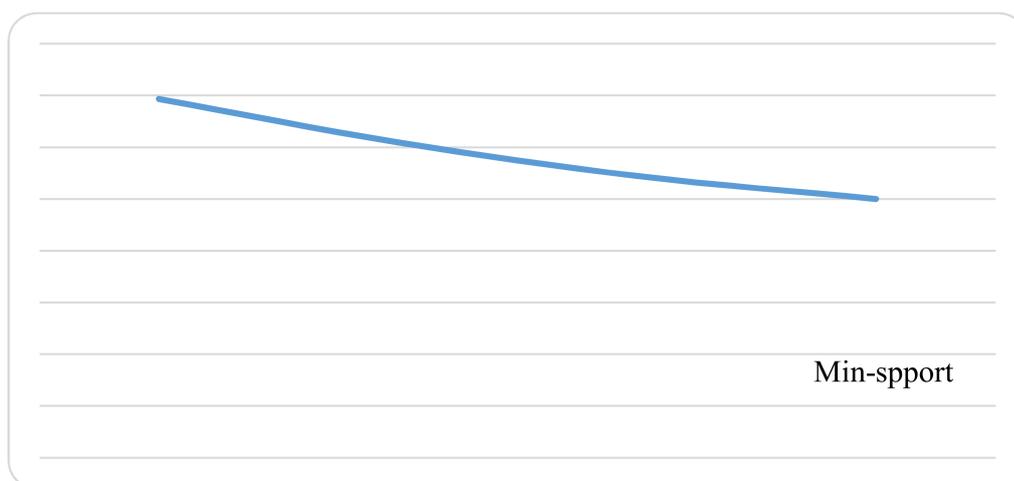


Figure 11. No. of neutrosophic association rules with different min-support threshold.

Experiment has been re-run using different min-support threshold values and the resulted neutrosophic association rules counts has been noted and listed in Table 11. Note the high values that are used for min-support threshold. Figure 12 illustrates the generated neutrosophic association rules for min-support threshold from 0.5 to 0.9.

Table 11. No. of neutrosophic rules with different min-support threshold.

Min-Support	0.5	0.6	0.7	0.8	0.9
1-itemset	11	9	9	6	5
2-itemset	50	33	30	11	10
3-itemset	122	64	50	10	10
4-itemset	175	71	45	5	5
5-itemset	151	45	21	1	1
6-itemset	88	38	8	0	0

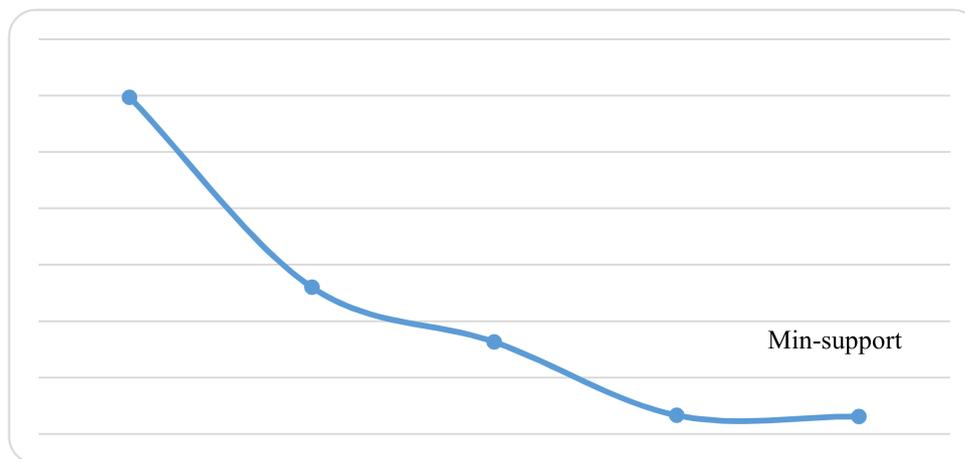


Figure 12. No. of neutrosophic rules for min-support threshold from 0.5 to 0.9.

Using the neutrosophic mining approach makes association rules exist for most of the min-support threshold domain, which may be sometimes misleading. We found that using the neutrosophic approach is useful in generating negative association rules beside positive association rules minings. Huge generated association rules provoke the need to re-mine generated rules (mining of mining association rules). Using suitable high min-support values may help in the neutrosophic mining process.

7. Conclusions and Future Work

Big data analysis will continue to grow in the next years. In order to efficiently and effectively deal with big data, we introduced in this research a new algorithm for mining big data using neutrosophic association rules. Converting quantitative attributes is the main key for generating such rules. Previously, it was performed by employing the fuzzy sets. However, due to fuzzy drawbacks, which we discussed in the introductory section, we preferred to use neutrosophic sets. Experimental results showed that the proposed approach generated an increase in the number of rules. In addition, the indeterminacy-membership function has been used to prevent losing rules from boundaries problems. The proposed model is more effective in processing negative association rules. By comparing it with the fuzzy association rules mining approaches, we conclude that the proposed model generates a larger number of positive and negative association rules, thus ensuring the construction of a real and efficient decision-making system. In the future, we plan to extend the comparison between the neutrosophic association rule mining and other interval fuzzy association rule minings. Furthermore, we seized the falsity-membership function capacity to generate negative association rules. Conjointly, we availed of the indeterminacy-membership function to prevent losing rules from boundaries problems. Many applications can emerge by adaptations of truth-membership function, indeterminacy-membership function, and falsity-membership function. Future work will benefit from the proposed model in generating negative association rules, or in increasing the quality of the generated association rules by using multiple support thresholds and multiple confidence thresholds for each membership function. The proposed model can be developed to mix positive association rules (represented in the truth-membership function) and negative association rules (represented in the falsity-membership function) in order to discover new association rules, and the indeterminacy-membership function can be put forth to help in the automatic adoption of support thresholds and confidence thresholds. Finally, yet importantly, we project to apply the proposed model in the medical field, due to its capability in effective diagnoses through discovering both positive and negative symptoms of a disease. All future big data challenges could be handled by combining neutrosophic sets with various techniques.

Author Contributions: All authors have contributed equally to this paper. The individual responsibilities and contribution of all authors can be described as follows: the idea of this whole paper was put forward by Mohamed Abdel-Basset and Mai Mohamed, Victor Chang completed the preparatory work of the paper. Florentin Smarandache analyzed the existing work. The revision and submission of this paper was completed by Mohamed Abdel-Basset.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gartner. Available online: <http://www.gartner.com/it-glossary/bigdata> (accessed on 3 December 2017).
2. Intel. Big Thinkers on Big Data (2012). Available online: <http://www.intel.com/content/www/us/en/bigdata/big-thinkers-on-big-data.html> (accessed on 3 December 2017).
3. Aggarwal, C.C.; Ashish, N.; Sheth, A. The internet of things: A survey from the data-centric perspective. In *Managing and Mining Sensor Data*; Springer: Berlin, Germany, 2013; pp. 383–428.
4. Parker, C. Unexpected challenges in large scale machine learning. In Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, Beijing, China, 12 August 2012; pp. 1–6.
5. Gopalkrishnan, V.; Steier, D.; Lewis, H.; Guszczka, J. Big data, big business: Bridging the gap. In Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, Beijing, China, 12 August 2012; pp. 7–11.
6. Chytil, M.; Hajek, P.; Havel, I. The GUHA method of automated hypotheses generation. *Computing* **1966**, *1*, 293–308.
7. Park, J.S.; Chen, M.-S.; Yu, P.S. An effective hash-based algorithm for mining association rules. *SIGMOD Rec.* **1995**, *24*, 175–186. [[CrossRef](#)]
8. Aggarwal, C.C.; Yu, P.S. Mining large itemsets for association rules. *IEEE Data Eng. Bull.* **1998**, *21*, 23–31.
9. Savasere, A.; Omiecinski, E.R.; Navathe, S.B. An efficient algorithm for mining association rules in large databases. In Proceedings of the 21th International Conference on Very Large Data Bases, Georgia Institute of Technology, Zurich, Switzerland, 11–15 September 1995.
10. Agrawal, R.; Srikant, R. Fast algorithms for mining association rules. In Proceedings of the 20th International Conference of Very Large Data Bases (VLDB), Santiago, Chile, 12–15 September 1994; pp. 487–499.
11. Hidber, C. Online association rule mining. In Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, Philadelphia, PA, USA, 31 May–3 June 1999; Volume 28.
12. Valtchev, P.; Hacene, M.R.; Missaoui, R. A generic scheme for the design of efficient on-line algorithms for lattices. In *Conceptual Structures for Knowledge Creation and Communication*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 282–295.
13. Verlinde, H.; de Cock, M.; Boute, R. Fuzzy versus quantitative association rules: A fair data-driven comparison. *IEEE Trans. Syst. Man Cybern. Part B* **2005**, *36*, 679–684. [[CrossRef](#)]
14. Huang, C.-H.; Lien, H.-L.; Wang, L.S.-L. An Empirical Case Study of Internet Usage on Student Performance Based on Fuzzy Association Rules. In Proceedings of the 3rd Multidisciplinary International Social Networks Conference on Social Informatics 2016 (Data Science 2016), Union, NJ, USA, 15–17 August 2016; p. 7.
15. Hui, Y.Y.; Choy, K.L.; Ho, G.T.; Lam, H. A fuzzy association Rule Mining framework for variables selection concerning the storage time of packaged food. In Proceedings of the 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Vancouver, BC, Canada, 24–29 July 2016; pp. 671–677.
16. Huang, T.C.-K. Discovery of fuzzy quantitative sequential patterns with multiple minimum supports and adjustable membership functions. *Inf. Sci.* **2013**, *222*, 126–146. [[CrossRef](#)]
17. Hong, T.-P.; Kuo, C.-S.; Chi, S.-C. Mining association rules from quantitative data. *Intell. Data Anal.* **1999**, *3*, 363–376. [[CrossRef](#)]
18. Pei, B.; Zhao, S.; Chen, H.; Zhou, X.; Chen, D. FARP: Mining fuzzy association rules from a probabilistic quantitative database. *Inf. Sci.* **2013**, *237*, 242–260. [[CrossRef](#)]
19. Siji, P.D.; Valarmathi, M.L. Enhanced Fuzzy Association Rule Mining Techniques for Prediction Analysis in Betathalassaemia's Patients. *Int. J. Eng. Res. Technol.* **2014**, *4*, 1–9.
20. Lee, Y.-C.; Hong, T.-P.; Wang, T.-C. Multi-level fuzzy mining with multiple minimum supports. *Expert Syst. Appl.* **2008**, *34*, 459–468. [[CrossRef](#)]

21. Chen, C.-H.; Hong, T.-P.; Li, Y. Fuzzy association rule mining with type-2 membership functions. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Bali, Indonesia, 23–25 March 2015; pp. 128–134.
22. Sheibani, R.; Ebrahimzadeh, A. An algorithm for mining fuzzy association rules. In Proceedings of the International Multi Conference of Engineers and Computer Scientists, Hong Kong, China, 19–21 March 2008.
23. Lee, Y.-C.; Hong, T.-P.; Lin, W.-Y. Mining fuzzy association rules with multiple minimum supports using maximum constraints. In *Knowledge-Based Intelligent Information and Engineering Systems*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 1283–1290.
24. Han, J.; Pei, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Elsevier: New York, NY, USA, 2011.
25. Hong, T.-P.; Lin, K.-Y.; Wang, S.-L. Fuzzy data mining for interesting generalized association rules. *Fuzzy Sets Syst.* **2003**, *138*, 255–269. [[CrossRef](#)]
26. Au, W.-H.; Chan, K.C. Mining fuzzy association rules in a bank-account database. *IEEE Trans. Fuzzy Syst.* **2003**, *11*, 238–248.
27. Dubois, D.; Prade, H.; Sudkamp, T. On the representation, measurement, and discovery of fuzzy associations. *IEEE Trans. Fuzzy Syst.* **2005**, *13*, 250–262. [[CrossRef](#)]
28. Smarandache, F. Neutrosophic set—a generalization of the intuitionistic fuzzy set. *J. Def. Resour. Manag.* **2010**, *1*, 107.
29. Abdel-Basset, M.; Mohamed, M. The Role of Single Valued Neutrosophic Sets and Rough Sets in Smart City: Imperfect and Incomplete Information Systems. *Measurement* **2018**, *124*, 47–55. [[CrossRef](#)]
30. Ye, J. A multicriteria decision-making method using aggregation operators for simplified neutrosophic sets. *J. Intell. Fuzzy Syst.* **2014**, *26*, 2459–2466.
31. Ye, J. Vector Similarity Measures of Simplified Neutrosophic Sets and Their Application in Multicriteria Decision Making. *Int. J. Fuzzy Syst.* **2014**, *16*, 204–210.
32. Hwang, C.-M.; Yang, M.-S.; Hung, W.-L.; Lee, M.-G. A similarity measure of intuitionistic fuzzy sets based on the Sugeno integral with its application to pattern recognition. *Inf. Sci.* **2012**, *189*, 93–109. [[CrossRef](#)]
33. Wang, H.; Smarandache, F.; Zhang, Y.; Sunderraman, R. Single valued neutrosophic sets. *Rev. Air Force Acad.* **2010**, *10*, 11–20.
34. Mondal, K.; Pramanik, S.; Giri, B.C. Role of Neutrosophic Logic in Data Mining. *New Trends Neutrosophic Theory Appl.* **2016**, *1*, 15.
35. Lu, A.; Ke, Y.; Cheng, J.; Ng, W. Mining vague association rules. In *Advances in Databases: Concepts, Systems and Applications*; Kotagiri, R., Krishna, P.R., Mohania, M., Nantajeewarawat, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4443, pp. 891–897.
36. Srinivas, K.; Rao, G.R.; Govardhan, A. Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques. In Proceedings of the 5th International Conference on Computer Science and Education (ICCSE), Hefei, China, 24–27 August 2010; pp. 1344–1349.

