Journal of
**Data science and Modern Techniques**

Research Article Open Access

# Why not MGK-measure Instead McNemar Measure?

**Chrisant Guyno Rasolonomenjanahary[1*], Ambeondahy[2] and André Totohasina[3]**

[1]Thematic doctoral school of the University of Toamasina, Madagascar

[2]ISSTM; University of Mahajanga, Madagascar

[3]ENSET; University of Antsiranana, Madagascar

[*]**Corresponding Author:** Chrisant Guyno Rasolonomenjanahary, Thematic doctoral school of the University of Toamasina, Madagascar, E-mail: r.guy.chris@gmail.com

**Received Date:** December 08, 2024 **Accepted Date:** January 08, 2025 **Published Date:** January 11, 2025

**Citation:** Chrisant Guyno Rasolonomenjanahary, Ambeondahy, André Totohasina (2025) Why not MGK-measure Instead McNemar Measure?. J Data Sci Mod Tech 2: 1-12

## Abstract

In the era of technologic advancement and artificial intelligence, a large amount of data is available. These data contain information, such as association rules, useful for decision-making. To extract them, we need information extraction tools such as association rule measures. To have reliable information, we must also use relevant measures, so we compared the association rule measures MGK and the McNemar measure. As a comparison result, we saw that the MGK measure is more accurate and more distinctive than the McNemar measure which confuses all forms of dependencies even independence.

**Keywords:** Information; Association Rules; Knowledge; McNemar; MGK Measure

## Introduction

We are currently in the age of advanced artificial intelligence that is, a machine able to react or to think like a human being [1]. In order to do this, it requires an enormous amount of information, both to train itself and to make decisions [2]. This information is obtained by analyzing a huge amount of data, or a database. There are several types of information in the literature, but what interests us most is information of the association rules type. Association rules is an information presented like $X \to Y$, that we can read as if X is present or true then Y will be present or true. This kind of information can help us to take a decision in more fields interacting with decision making even in medical health, social sciences, environment, more other [3]. In this article, we present what an association rule is and how its quality is measured. We then present and analyze two measures of association rules: the McNemar and $M_{GK}$.

### Association Rules and their Measurements

An association rule is a pair (X,Y) noted $X \to Y$ where *X* and *Y* are disjoint patterns (or conjunctions of binary variables). An association rule of type $X \to Y$ association rule takes the form "If condition, then result". It has a premise part (or antecedent) composed of a set of variables *X* and a conclusion part (or consequent) composed of a set of variables *Y* disjoint from *X*. Such a rule is used to discover whether transactions that verify pattern *X* tend to verify pattern *Y* as well. For this reason, it is extracted from a formal database, as defined below. An association rule is entirely characterized by its contingency table, which is the basis for calculating association rule evaluation measures.

### Definition 2.1: Formal Database

Let $A = \{x_1, x_2, \ldots, x_m\}$ a set of m variable items or attributes, and $T = \{t_1, t_2, \ldots, t_n\}$ a set of n transactions or entities defined on the set of attributes. A subset $X_i \in P(A)(P(A)$ is the set of parts of *A*) is called a pattern and $\overline{X}_i \subset A$ its complement. Each transaction $t_i$ consists of forming a subset of attributes of $A, X_k$. A transaction is said to $t_i$ satisfies $x_k$ if the transaction $t_i$ contains $x_k$ in which case we write $t_i[x_k]=1$ otherwise $t_i[x_k]=0$. A transaction contains a pattern if all the attributes that make up the pattern are contained in the transaction. The triple ( *T,A,R*), where *T* is the set of transactions or objects, *A* the set of attributes or variables and *R* a binary relationship from *T* to *A* [7].

### Definition 2.2: Pattern extension

Let *K* = (*T,A,R*) be a binary data mining context. The extension of the pattern *X* denoted $X'$ the set of *t* of *T* such that for all *x* from *X*, *tRx*:

$$X' = \{t \in T | \forall x \in X, tRx\}.$$

In the set of transactions, which we have denoted *T* we can define a discrete probabilized space (*T,P(T),P*) where *P* is the discrete uniform probability [4].

In health study, as used in “Improved extension of MGK in several premises simplified applied on COVID19 study” [5], the motif *X* can represent the symptoms, *Y* a disease and $X'$ is a list of all patients presenting the symptoms *X*.

So, if $X' \subset T then P(X') = \frac{|X'|}{|T|} = \frac{n_X}{n}$ . If *X* and *Y* being two patterns such $P(X' \cap Y') \geqslant \alpha$ where α is a predefined value indicating the significance of the combination (*X,Y*). From these significant combinations, we can build an association rule such as $X \to Y$ or $Y \to X$.

### Definition 2.3: Formal definition of an association rule

An association rule of a formal context *K* = (*T,A,R*) is a pair (*X,Y*) noted $X \to Y$ where *X* and *Y* are patterns and $Y \neq \phi$ is commonly read as "if *X*, then *Y*". It expresses an association or oriented link between X and Y, and X is said to be the premise and Y the consequent.

According to these definitions, an association rule is generated from a large database, which we have called a formal database. As a result, many association rules can be generated, but not all of them can be considered significant. To assess their significance, we need to use a reliable sorting tool, known as a probabilistic measure of interest. At present, there are several such measures, for example: Confidence, Lift, positive predictive values (PPV), negative predictive values (NPV), Odd-Ratio, the McNemar test, etc., around one hundred and nine (109), in the literature. The large number of measures is due to the fact that researchers are striving to improve the measures in order to have confident and reliable association rules. A study leaded by Bruce [6] has shown the shortcomings of the Odd-Ratio test compared with the $M_{GK}$ test. This means that we try to find a good measure that can give us the best result. In the reminder of this article, we will compare the McNemar measure with the $M_{GK}$ measure, which has already overtaken the Odd-Ratio measure. However, a measure must respect some forms and conditions.

**Definition 2.4 (A quality measure: μ).**

Let $X \in P(A) and Y \in P(A)$ be patterns. A probabilistic quality measure is a real function μ of $P(A) \times P(A)$ such that for any association rule $X \rightarrow Y, \mu(X \rightarrow Y)$ is a real value calculated from the four quantities: $nP(X'), P(Y')\ and\ P(X' \cup Y')$, where $P$ denotes the discrete uniform probability on the probability space $(A,P(A))$.

**Property 2.1 (Measure property).**

1. A measure is said to be symmetrical if $\mu(X \rightarrow Y) = \mu(Y \rightarrow X)$.

2. A measure is said to be oriented if there is at least one association rule $X \rightarrow Y$ we have $\mu(X \rightarrow Y) \neq \mu(Y \rightarrow X)$.

Several researchers have focused on the criteria for evaluating a good quality measure (André Totohasina in 2008 [7], Lenca Philippe in 2011 [8], Grissa in 2013 [9], Lan Phuong in 2016 [10] and Rakotomalala in 2019 [11], among these works, we were able to select some criteria:

- Measurement comprehensibility for the user (interpretable);
- Nature of the rules targeted by the measure;
- Sensitivity to the appearance of examples and counter-examples;
- Direction of measurement variation;
- Type of variation: linear/non-linear;
- Impact of the scarcity of the consequent;
- Sensitivity to data size;
- Discriminant character of the measurement;
- Use of a pruning weir;
- Measurement-induced classification
- Contextual behavior of the rules studied;
- Deviation from equilibrium;
- Contradiction of the user's priori knowledge;
- Noise sensitivity.

According to the literature, measures that meet most of these criteria are ranked better than others. In what follows, we will not compare them on the basis of these criteria, but rather on their mathematical expressions.

**McNemar's Measurement**

The McNemar measure, generally used in the McNemar test, is often used to evaluate the dependence of two binary qualitative variables, or of the same variable evaluated under two different conditions or instants [12]. In her article entitled: "Matched analysis for paired binary data (McNemar test)" published in the American Journal of Orthodontics and Dentofacial Orthopedics (AJO-DO), Despina Koletsi presented an application of the McNemar test to the identification of early root resorption of lateral incisors [13]. For his part, Jean Claude Regnier used the McNemar test to assess the dependence of students' responses to exercises [14].

**Definition 3.1: Definition of the McNemar test**

**Consider the following contingence table (Table 31):**

| Test 2: Y | | | | |
|---|---|---|---|---|
| | | 0 | 1 | |
| Test 1: X | 0 | $n_{\bar{X}\bar{Y}}$ | $n_{\bar{X}Y}$ | $n_{\bar{X}}$ |
| | 1 | $n_{X\bar{Y}}$ | $n_{XY}$ | $n_X$ |
| | | $n_{\bar{Y}}$ | $n_Y$ | $n$ |

**Table 3-1:** Contingency table

**Interpreting table values in Table 31:**

$n_{\overline{XY}}$: number of individuals who answered 0 (or absence of X) in test 1 and 0 (or absence of Y) in test 2

$nXY$: number of individuals who answered 1 (or presence of X) to test 1 and then 1 (or presence of Y) to test 2

$n_{\overline{X}Y}$: number of individuals who answered 0 (or absence of X) to test 1 then 1 (or presence of Y) to test 2

$n_{X\overline{Y}}$: number of individuals who answered 1 (or presence of X) to test 1 and 0 (or absence of Y) to test 2

McNemar's measure is defined by [12].

$$\chi^2_{\mathrm{Mc}}(X \to Y) = \frac{(n_{X\bar{Y}} - n_{\bar{X}Y})^2}{n_{X\bar{Y}} + n_{\bar{X}Y}}$$

This expression follows a chi-square distribution with one degree of freedom and α risk threshold. Remember that in the definition of a measure (definition 2.4), it should be expressed in terms of four elements: $n, P(X'), P(Y')\ and\ P(X' \cup Y')$.

So, let's re-express McNemar's measure in accordance with the indications for formulating the measure. We then have the following proposition.

**Proposition 3.1:**

Let *X* and *Y* be two patterns, P() the uniform probability. The McNemar measure applied to the association rule $X \to Y$ is defined by:

$$\chi^2_{Mc}(X \to Y) = \frac{(P(X') - P(Y'))^2}{P(X') + P(Y') - 2P(X' \cap Y')} \times n$$

**Proof**

Given that X and Y are patterns in a formal database *K = (T,A,R)*, we can define a probabilized space on the formal database K with uniform probability, such that $P(X') = \frac{n_X}{n}$.

Multiply the numerator by $\frac{n^2}{n^2}$ and the denominator by $\frac{n}{n}$ then we have:

$$\chi^2_{\mathrm{Mc}}(\mathrm{XY}) = \frac{\frac{n^2}{n^2}\left(n_{X\overline{Y}} - n_{\overline{X}Y}\right)^2}{\frac{n}{n}\left(n_{X\overline{Y}} + n_{\overline{X}Y}\right)}$$

$$= \frac{n^2\left(\frac{n_{X\overline{Y}}}{n} - \frac{n_{\overline{X}Y}}{n}\right)^2}{n\left(\frac{n_{X\overline{Y}}}{n} + \frac{n_{\overline{X}Y}}{n}\right)}$$

$$= \frac{n\left(P(X \cap \overline{Y}) - P(\overline{X} \cap Y)\right)^2}{P(X \cap \overline{Y}) + P(\overline{X} \cap Y)}$$

We know that

$$\mathrm{P}\left(\mathrm{X} \cap \overline{Y}\right) = \mathrm{P(X)} \text{ - } \mathrm{P}\left(X \cap \mathrm{Y}\right)$$

The same goes for

$$\mathrm{P}\left(Y \cap \overline{X}\right) = \mathrm{P(Y)} \text{ - } \mathrm{P}\left(X \cap \mathrm{Y}\right)$$

We then have

$$\chi^2_{\mathrm{Mc}}\left(X \to Y\right) = \frac{\mathrm{n} \times \left(\mathrm{P}\left(\mathrm{X}\right) - \mathrm{P}\left(\mathrm{Y}\right)\right)^2}{P\left(X\right) + P\left(Y\right) - 2P(X \cap Y)}.$$

Note: In this new expression of the McNemar measure $X^2_{MC}(X \to Y) = X^2_{MC}(Y \to X)$ the measure is therefore symmetrical.

A McNemar test validation use the chi-square table with one degree of freedom.

**MGK Measurement**

The MGK interest measure of association rules goes by various names independently, depending on the researcher and the year of its discovery: inspired by the Loevinger index, MGK (Guillaume-Kenchaff measure) was independently proposed and named in 2000 by Guillaume" ION (Implication Oriented Normalized) in 2003 by Totohasina, CPIR (Conditionnal Probability Incrementation Ratio) in 2004 by Wu and Zhang, verifying the oriented implicative property of Brin and his team in 1997 [15]. Due to the expression of a minimum condition and efficiency ratio for extracting non-redundant rules, this measure is both more accurate and comprehensible. By definition, the MGK measure is expressed as:

$$\mathrm{M_{GK}}\left(\mathrm{X} \to \mathrm{Y}\right) = \begin{cases} \frac{\mathrm{P(Y'—X')} - \mathrm{P(Y')}}{1 - \mathrm{P(Y')}}, \mathrm{SiXfavorY} \\ \frac{\mathrm{P(Y'—X')} - \mathrm{P(Y')}}{\mathrm{P(Y')}}, \mathrm{siXdisfavorY} \end{cases}$$

**Indeed**

- if X favors Y, on the one hand

$$\mathrm{P(Y'|X')} \geqslant \mathrm{P}\left(\mathrm{Y'}\right)$$

so

$$P(Y'|X') - P(Y') \geqslant 0$$

on the other hand

$$1 \geqslant P(Y'|X') \geqslant 0$$

so

$$1 - P(Y') \geqslant P(Y'|X') - P(Y') \geqslant -P(Y')$$

so

$$1 - P(Y') \geqslant P(Y'|X') - P(Y') \geqslant 0 (because\ P(Y'|X') - P(Y') \geqslant 0)$$

So

$$1 \geqslant 1 - P(Y') \geqslant 0$$

- if X disfavors Y, on the one hand

$$P(Y'|X') \leqslant P(Y')$$

So

$$P(Y'—X') - -P(Y') \leqslant 0$$

and on the other hand

$$1 \geqslant P(Y'|X') \geqslant 0$$

so

$$1 - P(Y') \geqslant P(Y'|X') - P(Y') \geqslant -P(Y')$$

so

$$0 \geqslant P(Y'|X') - P(Y') \geqslant -P(Y'), (carP(Yt|Xt) - P(Yt) \leqslant 0)$$

so

$$0 \geqslant P(Y'|X') - P(Y') \geqslant -1.$$

If we denote by $M^{f}_{GK}(X \rightarrow Y) = \frac{P(Y'—X') - P(Y')}{1 - P(Y')}$ called the favorable component, and $M^{d}_{GK}(X \rightarrow Y) = \frac{P(Y'—X') - P(Y')}{P(Y')}$ called the unfavorable component, then the measure MGK can be expressed

$$M_{GK}(X \rightarrow Y) = \begin{cases} M^{f}_{GK}(X \rightarrow Y), siXfavoriseY \\ M^{d}_{GK}(X \rightarrow Y), siXdé\ favoriseY \end{cases}$$

**Measurement Reference Situations $M_{GK}$**

A definition of reference situations is a criterion justifying the quality of a measurement. The MGK measure presents these situations [7,11] as follows:

Incompatibility: X and Y are incompatible if and only if $\mathrm{M_{GK}}\,(\mathrm{X} \rightarrow \mathrm{Y}) = -1$.

Unfavorable situation or negative dependency: X disfavor Y if and only if $-1 < \mathbf{M_{GK}}\,(\mathbf{X} \rightarrow \mathbf{Y}) < 0$.

Situation of independence: X and Y are independent if and only if $\mathbf{M_{GK}}\,(\mathbf{X} \rightarrow \mathbf{Y}) = 0$

Favorable situation or positive dependence: X favor Y if and only if $0 < M_{GK}(X \rightarrow Y) < 1$.

Logical implication situation: X logically implies Y if and only if $M_{GK}(X \rightarrow Y) = 1$

Equilibrium situation: in an equilibrium situation, i.e. $|X \cap Y| = |X \cap Y|$, the measurement $M_{GK} = \pm 1/2$

**Measurement validation threshold $M_{GK}$**

Like all the measures we have seen above, the $M_{GK}$ measure has a slightly different validation threshold: it is calculated using the Chi-square statistic, based on the principle of assigning confidence to the rules that will be generated. The $M_{GK}$ threshold is expressed as:

$$M_{\mathrm{GK}}\mathrm{threshold}(\alpha) = \pm\sqrt{\frac{1}{n} \cdot \frac{n_{\overline{X}}}{n_X} \cdot \frac{n_Y}{n_{\overline{Y}}}\chi^2(\alpha)}$$

An association rule is therefore valid, according to the measure $M_{GK}$ if the absolute value of its $M_{GK}$ is greater than the absolute value of the threshold $M_{GK}$ threshold (α) and this, at a confidence level 1- α.

**Comparative Analysis of McNemar and $M_{GK}$**

By definitions,

$$\chi^2_{\mathrm{Mc}}\,(\mathrm{X} \rightarrow \mathrm{Y}) = \frac{(P(X') - P(Y'))^2}{P(X') + P(Y') - 2P(X' \cap Y')} \times n$$

And

$$\mathrm{M_{GK}}\,(\mathrm{X} \rightarrow \mathrm{Y}) = \begin{cases} \frac{\mathrm{P(Y'—X')-P(Y')}}{\mathrm{1-P(Y')}}, \mathrm{SiXfavorY} \\ \frac{\mathrm{P(Y'—X')-P(Y')}}{\mathrm{P(Y')}}, \mathrm{siXdisfavorY} \end{cases}.$$

As the both measures are expressed with $n, P(X'), P(Y')\ and\ P(X' \cup Y')$, we can express each with other it means

$\chi^2_{\mathrm{Mc}}\,(\mathrm{X} \rightarrow \mathrm{Y}) = f\,(\mathrm{M_{GK}}\,(\mathrm{X} \rightarrow \mathrm{Y}))\ or\ M_{GK}(X \rightarrow Y) = f(\chi^2_{MC}(X \rightarrow Y))$

Considering, $\chi^2_{\mathrm{Mc}}\,(\mathrm{X} \rightarrow \mathrm{Y}) = f\,(\mathrm{M_{GK}}\,(\mathrm{X} \rightarrow \mathrm{Y}))$, we have the following relation:

$$\chi^2_{MC}\,(X \rightarrow Y) = \frac{n \times (P\,(X) - P\,(Y))^2}{P\,(Y) + P\,(X)\,(1 + 2M_{GK}\,(X \rightarrow Y) + P\,(Y)\,(1 + 2M_{GK}\,(X \rightarrow Y)))}.$$

We can see that we don't have the usual relationship, so we will take a few examples already used by Totohasina to compare the Chi-square measure with the measure $M_{GK}$ in the five reference situations (positive dependence (Table 51), negative dependence (Table 52), independence (Table 53), incompatibility (Table 54) and logical implication (Table 55)) [7].

| $X \backslash Y$ | $Y$ | $\bar{Y}$ | |
|---|---|---|---|
| $X$ | 3000 | 2000 | 5000 |
| $\bar{X}$ | 2500 | 2500 | 5000 |
| | 5500 | 4500 | 10000 |

| Measure | Value |
|---|---|
| $\chi^2$ | 101 |
| $\chi^2_{Mc}$ | 250 |
| $M_{GK}$ | 0.11 |

**Table 5-1:** Positive dependence $\chi^2 = 101, \mathrm{M_{GK}} = +0.11, \chi^2_{\mathrm{Mc}} = 250$

| $X \backslash Y$ | $Y$ | $\bar{Y}$ | |
|---|---|---|---|
| $X$ | 1000 | 3000 | 4000 |
| $\bar{X}$ | 4500 | 1500 | 6000 |
| | 5500 | 4500 | 10000 |

| Measure | Value |
|---|---|
| $\chi^2$ | 2424 |
| $\chi^2_{Mc}$ | 300 |
| $M_{GK}$ | -0.54 |

**Table 5-2:** Negative dependence $\chi^2 = 2424, M_{GK} = -0.54, \chi^2_{MC} = 300$

| $X \backslash Y$ | $Y$ | $\bar{Y}$ | |
|---|---|---|---|
| $X$ | 2200 | 1800 | 4000 |
| $\bar{X}$ | 3300 | 2700 | 6000 |
| | 5500 | 4500 | 10000 |

| Measure | Value |
|---|---|
| $\chi^2$ | 0 |
| $\chi^2_{Mc}$ | 441.17 |
| $M_{GK}$ | 0 |

**Table 5-3:** Independence $\chi^2 = 0, M_{GK} = 0, \chi^2_{MC} = 444.17$

| $X \backslash Y$ | $Y$ | $\bar{Y}$ | |
|---|---|---|---|
| $X$ | 0 | 2000 | 2000 |
| $\bar{X}$ | 6000 | 2000 | 8000 |
| | 6000 | 4000 | 10000 |

| Measure | Value |
|---|---|
| $\chi^2$ | 3750 |
| $\chi^2_{Mc}$ | 2000 |
| $M_{GK}$ | -1 |

**Table 5-4:** Incompatibility $\chi^2 = 3750, M_{GK} = -1, \chi^2_{MC} = 2000$

| $X$ \ $Y$ | $Y$ | $\bar{Y}$ | |
|---|---|---|---|
| $X$ | 3000 | 0 | 3000 |
| $\bar{X}$ | 3000 | 4000 | 7000 |
| | 6000 | 4000 | 10000 |

| Measure | Value |
|---|---|
| $\chi^2$ | 2857 |
| $\chi^2_{Mc}$ | 3000 |
| $M_{GK}$ | 1 |

**Table 5-5:** Logical implication $\chi^2 = 2857, M_{GK} = 1, \chi^2_{MC} = 3000$

| Measure | Positive dependence | Negative dependency | Independence | Incompatibility | Logical implication |
|---|---|---|---|---|---|
| $\chi^2$ | 101 | 2424 | 0 | 3750 | 2857 |
| $\chi^2_{Mc}$ | 250 | 300 | 441.17 | 2000 | 3000 |
| $M_{GK}$ | 0.11 | -0.54 | 0 | -1 | 1 |
| $M_{GK}$threshold(0.05) | 0.02167 | 0.02654 | 0.02654 | 0.04801 | 0.03667 |

**Table 5-6:** Summary of Chi-square, McNemar and MGK values for the five reference situations for the examples considered

We notice that at the last table (Table 56), both measures $\left(X^2, X^2_{MC}\right)$ are always positive, belong to the interval $[0; +\infty]$, which means that the measurements are symmetrical, i.e. $X \rightarrow Y$ is equivalent to $Y \rightarrow X$. At independence, the McNemar measure is still positive and different from 0. Compared with its value at negative dependence, we can interpret this as a strong link, since at independence its value is higher than at negative dependence. These facts mean that the Chi-square and McNemar measures are unreliable and fail to meet many of the characteristics of a quality measure of association rules. However, the measure $M_{GK}$ measures differentiate situations well, moreover its values are only between 0 and 1, differentiated by signs; so can be interpreted as a probability such that the signs only indicate the direction of the link: negative if opposite, i.e. the presence of the other implies the absence of the other, and positive if the same orientation. To sum up, the Chi-square and McNemar measures arse unreliable and very confusing, which is not the case for the measure $M_{GK}$.

For a risk α = 0.05, the χ2 (α) with one degree of freedom is equal to 3.841, then:

in positive dependence (Table 5-1), $M_{GK}$ threshold (0.05) = 0.02654 and $M_{GK}$ = 0.11 then we accept the implication X→Y (because $M_{GK} \geq M_{GK}$ threshold)

in negative dependence (Table 5-2), $M_{GK}$ threshold (0.05) = -0.02654 and $M_{GK}$ = -0.54 , as the both values are negative, and $M_{GK}$ threshold ≥ $M_{GK}$ then one of $\bar{X} \rightarrow Y$ and $X \rightarrow \bar{Y}$ is valid

- in independence (Table 5-3), $M_{GK}$ threshold (0.05) = 0.02654 and $M_{GK}$ = 0 then we cannot say anything

in incompatibility (Table 5-4), $M_{GK}$ threshold (0.05) = -0.04801 and $M_{GK}$ = -1, then certainly one of →Y and X→ $\bar{Y}$

- in logical implication (Table 5-5) $M_{GK}$ threshold (0.05) = 0.03667 and $M_{GK}$ = 1 so we accept without a doubt the implication X→Y

## Conclusion

Extracting knowledge from a database is a crucial step in the development of artificial intelligence, since the information is used to make the right decisions. This information includes association rules, which are extracted from a formal database. The quality of an association rule is measured using a probabilistic function, with at least 106 measures listed in the literature. Even if these measures are used to measure the quality of an association rule, their ability to validate rules must also be studied. Therefore, we analyzed and observed the behavior of two measures: the McNemar measure, often used in medicine, and the $M_{GK}$. As a result, we found that the McNemar measure is very confusing and indiscriminately validates association rules in the five reference situations, on the one hand. The $M_{GK}$ measure, on the other hand, distinguishes between cases such as -1 for incompatibility, a value between -1 and 0 for negative dependency, 0 for independence, between 0 and 1 for positive dependency and 1 for logical implication. So, it is best to use the $M_{GK}$ measure, thanks to its precise filters and reliability.

## Conflict of Interest

The authors declare no potential conflict of interests.

## References


1. BERTOLUCCI M (2023) Artificial intelligence in the public sector: literature review and research agenda. Management & Public Management Review, 12: 118-39.

2. Andrew L, Steve L (2021) Artificial intelligence foundations : learning from experience. BCS The Chaetered Institute For It.

3. Dopico JR, Calle JD, Sierra AP (2009) Encyclopedia of Artificial Intelligence. Information Science Reference.

4. Totohasina A (2008) Contribution to the study of quality measures of association rules: normalization under five constraints and MGK cases.

5. Rasolonomenjanahary CG, Ambeondahy, Totohasina A (2023) Improved Extension of MGK on Several Premisses Simplified. International Journal of Innovative Science and Research Technology, 2171-80.

6. Bemena BM, Feno DR, Totohasina A (2022) The Effectiveness of the MGK Measure against the Odds Ratio in the Epidemiological Study. International Journal of Mathematics and Mathematical Sciences, 12.

7. Totohasina A, Rajaonasy FD (2008) The quality of association rules: comparative study of MGK and Confidence measures. African Conference on Research in Computer Science and Applied Mathematics, 561-8. Morocco.

8. Philippe L, Stéphane L (2011) The choice of a good quality measure, a condition for the success of a data mining process. Data mining workshop, applications, case studies and success stories (associated with the French-speaking international conference on knowledge extraction and management).

9. Grissa D (2013) Behavioral study of knowledge extraction interest measures. Blaise Pascal University.

10. Phan LP, Phan NQ, Phan VC, Huynh HH, Huynh HX, Guillet F (2016) Classification of objective interstingness measures. EAI Endorsed Transactions on Context-aware Systems and Applications.

11. Rakotomalala HF (2019) Implicative and Cohesive Hierarchical Classification according to the MGK measurement - application in computer science teaching.

12. McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. psychometrika, 153-7.

13. Despina K, Nikolaos P (2017) Matched analysis for paired binary data (Mc Nemar test). American journal of Orthodontics and Dentofacial Orthopedics, 222-3.

14. Jean-Claude R (2016) Mac Nemar Test and Implicative Statistical Analysis. RNTI, 271-7.

15. Brin S, Motwani R, Silverstein C (1997) Beyond market basket: Generalizing Association rules to correlation. ACM SIGMOD Record, 265-76.