International Journal of Applied Data Science
Volume 2, Issue 1, 2026

EQUIVALENT MODELS IN ASSOCIATION RULES ANALYSIS

Zaini[1],Yusof Hussain[2]
Fakulti Sains Sosial dan Kemanusiaan Universiti Kebangsaan Malaysia.

Article Info

Received: 12-09-2025 **Revised:03-10-2025** **Accepted:17-01-2026** **Published:10-02-2026**

Abstract: A set of rules may be extracted from any dataset using association rules analysis, a potent tool for examining associations between elements. However, it is not often understood that many sets of rules, or comparable models, may describe a same dataset. In actuality, since they presume the model is unique, the majority of research on the model's quality might be deceptive. The literature has not yet examined these phenomena. In this paper, we show that there are comparable models for every dataset and provide a way to transform any model into its dominant model, which is then suggested as the benchmark model. We also describe how statistical model selection and decision tree analysis are impacted by the phenomena of equivalent models. It is shown that by lowering the decision rules to the dominant model, decision tree analysis choice rules may always be made simpler. For demonstration, both actual and simulated datasets are used.

Keywords*:* Association rules analysis, Decision tree analysis, Dominant model, Equivalent models.

Introduction

A well-liked data-mining method is association rules analysis, which was first presented in the early 1990s (Agrawal and Srikant 1994). It is a technique for investigating the connections between objects (variables) in terms of rules. Finding the rules that best describe a certain dataset is the core principle. These regulations need to fulfill certain requirements, such as a certain level of trust and assistance. Biology (Becquet et al. 2002; Kuo et al. 2007), business and marketing (Wang et al. 2005; Shih et al. 2010), geography (Appice et al. 2003; Lee et al. 2007), agriculture (Matsumoto 1998; Mclver and Friedl 2002), education (Garcia et al. 2007; Merceron and Yacef 2008), photography (Tesic 2003; Liu et al. 2008), and economics (Dopfer and Potts 2004; Mosson et al. 2008). Numerous developments have been made, such as techniques for choosing sets of rules to represent datasets (Bayardo 1997; Liu et al. 1998; Dong et al. 1999; Bastide et al. 2000; Yin and Han 2003; Webb and Zhang 2005; Li 2006; Thabtah 2007; Grudzinski et al. 2010).

However, there is a basic occurrence that has been overlooked, which might result in a comparison of rules selection approaches that is not suitable. This phenomena is the result of
"equivalent models," in which two or more association rules models may accurately reflect a dataset.

A dataset is created using a known or provided model in the literature (Etchells and Lisboa 2006; Guerreiro and Trigueiros 2010; Marghny and El-Semman 2005); that is, the produced dataset is used to evaluate each approach under review. The only rules deemed to be accurate are those found in the model. A approach is deemed competent if it discovers the same rules as the provided model. However, a technique is deemed incompetent if it discovers rules that deviate from the model.

However, as the following example shows, such a comparison may not be accurate. Assume that the model, which consists of two rules, is used to create the dataset in question:

Rule 1: Y equals 1 if X1 = 1.

Rule 2: Y = 1 if X2 = 1, X3 = 1, and X4 = 1.

Examine the two models (U and V) that follow:
Model U
Rule U1: Y equals 1 if X1 = 1.
Rule U2: Y = 1 if X2 = 1, X3 = 1, and X4 = 1.
Model V
Rule V1: Y = 1 if X1 = 1 and X3 = 0. Rule V2: Y = 1 if X1 = 1 and X4 = 1.
Rule V3: Y = 1 if X1 = 1, X2 = 1, and X3 = 1.
Rule V4: Y = 1 if X1 = 0, X2 = 1, X3 = 1, and X4 = 1. Rule V5: Y = 1 if X1 = 1, X2 = 0, X3 = 1, and X4 = 0.
Both Model U and Model V provide the same answers. The two models are equal, to put it another way. Comparisons across approaches and criteria may be deceptive if the equivalency of models is not completely appreciated. The literature has not yet examined this phenomena.
If there are many models available, the next problem is to describe the dataset or the connections between variables. Our main conclusion on this topic is that there is only one distinct "dominant model," even if several models may reflect the same information. In order to describe the connections between variables across all the comparable models, we suggest using the dominant model as the standard. We provide a clear and consistent explanation of the connections between the variables by applying the dominant model. To put it another way, the dominant model outperforms the other similar models in explaining the data while maintaining the same capacity for prediction. Inappropriate comparisons between the rules selection approaches may be avoided by using the dominant model to confirm equivalence between the provided models. To put it another way, each model has a dominating model. Two or more models are comparable if they share the same dominant model.
Analysis of association rules is not the only use of comparable models. Specifically, decision tree analysis is similarly affected by analogous models. An examination of decision trees is defined by a hierarchical structure that makes it possible to reduce any set of decision rules to a smaller set of rules. Additionally, statistical methods that use variables produced by decision tree analysis to create a logistic regression model are likewise subject to the influence of analogous models. We demonstrate that by limiting the decision rules to the dominant model, it is always possible to simplify the decision rules from decision tree analysis and the resulting variables.
Investigating this equivalence phenomena is the aim here. The definitions, the dominant model, the verification, and the theoretical characteristics of analogous models are all covered in our discussion of the key concepts and analysis. We demonstrate how comparable models directly impact statistical model selection techniques, decision tree analysis, and association rule analysis.
The structure of this document is as follows. In Section 2, an example is used to explain the fundamental idea of comparable models. The dominant model for analogous models is suggested in Section 3, along with the method for deriving the dominant model. The theoretical characteristics of the corresponding models are explained in Section 4. The connection between this phenomena and the popular MONK's dataset is shown in Section 5. The impact of comparable models with practical applications is shown in Section 6. The parallel effect of comparable models on decision tree analysis is explained in Section 7, along with how this phenomenon impacts practical applications. A discussion and closing thoughts are provided in Section 8.

An illustrative example

First, we illustrate the equivalency phenomenon of association rules analysis through an example showing that multiple models can be used to represent the same dataset. Note that this phenomenon is very common but has never been acknowledged in literature.
Consider an example with five binary variables (X1–X5) and a binary response (Y) with classes 0 and 1. Model A, which consists of two rules, is used to generate response Y:
Rule A1: If X1 = 1, then Y = 1.
Rule A2: If X2 = 1 and X3 = 1, then Y = 1.
The generated dataset can be used to compare the methodologies in terms of the respective performance of each. For this dataset, any technique that finds both Rule A1 and Rule A2 is considered competent. Now consider Model B, which consists of 4 rules:
Rule B1: If X1 = 1 and X2 = 1, then Y = 1. Rule B2: If X1 = 1 and X3 = 1, then Y = 1.

Rule B3: If X1 = 1, X2 = 0, and X3 = 0, then Y = 1. Rule B4: If X1 = 0, X2 = 1, and X3 = 1, then Y = 1.
Based on the current prevailing methodology, Model B would be considered incompetent because it did not find any rule in the given model (Model A). However, Model B generates responses that are identical to those produced by Model A. In other words, these two models are "equivalent." A lack of knowledge about model equivalence is likely to result in misleading comparisons between methods and rules selection criteria. Note that all the models shown in
Table 1 will generate responses that are identical to those generated by Model A and Model B, such that all of these models are equivalent.

Table 1: Illustrated Equivalent Models

Model	Rules
C	Rule C1: If $X_1 = 0$ and $X_2 = 0$, then Y = 0. Rule C2: If $X_1 = 0$ and $X_3 = 0$, then Y = 0.
D	Rule D1: If $X_1 = 1$, then Y = 1. Rule D2: If $X_1 = 0$, $X_2 = 1$, and $X_3 = 1$, then Y = 1.
E	Rule E1: If $X_1 = 0$ and $X_2 = 0$, then Y = 0. Rule E2: If $X_1 = 0$, $X_2 = 1$, and $X_3 = 0$, then Y = 0.
F	Rule F1: If $X_1 = 1$ and $X_4 = 0$, then Y = 1. Rule F2: If $X_1 = 1$ and $X_4 = 1$, then Y = 1. Rule F3: If $X_2 = 1$ and $X_3 = 1$, then Y = 1.

It is evident that given differences between the terms and the number of rules between these equivalents models (Models A–F), that each model explains the relationships among the variables differently. This becomes another issue associated with equivalent models. Although it is possible, as shown, for a large number of models to represent the same dataset, there is only one unique dominant model to which all the equivalent models can refer. In fact, the dominant model serves as a point of reference whereby the explanation of the relationships among the variables will be consistent and concise. The dominant model can also be used to verify equivalency among models in order to eliminate unfair comparisons among the rules selection methodologies and criteria. The next section presents a thorough study of equivalent models including important definitions related to model equivalency. The concept of a dominant model as a representation of equivalent models is proposed, and the procedure for converting equivalent models into the dominant model is described.

Equivalent Models and the Dominant Model

This section provides a method for converting any given model into its dominant model. First, we provide definitions of all the significant terms used. Then, we propose a method for converting any given models into the dominant model. Note that the dominant model can also be used to verify equivalency among models, as models that have the same dominant model are all equivalent.
Definitions
Definition 1: Models are called equivalent if they generate identical responses.
Note that Models A–F in Section 2 are all equivalent. For simplicity, our study separates the equivalent models into two classes of response: class 0 and class 1. The equivalent model with all the rules giving conditions for Y = 0 is called the equivalent model for class 0. Likewise, the

equivalent model with all the rules giving conditions for Y = 1 is called the equivalent model for class 1. Note that equivalent models from one class can be viewed as complementing the equivalent models of the other class, as will be discussed later.

Definition 2: Of all the equivalent models, the dominant model is the model with the fewest condition terms.

A condition term is represented by one variable and its level, e.g., Xi = 1. We count all the repeated condition terms. As there are two classes for a binary response, there are two dominant models—one dominant model for each class. Note that the dominant model for each class is unique. The proof is provided in Section 4.

Consider the example in Section 2: Model A is the dominant model for class 1, and Model C is the dominant model for class 0. For Model A, the total number of condition terms is three: one term from Rule A1 (X1 = 1), and two terms from Rule A2 (X2 = 1 and X3 = 1). For Model C, the total number of condition terms is four: two terms from Rule C1 (X1 = 0 and X2 = 0), and two terms from Rule C2 (X1 = 0 and X3 = 0). Note that the dominant model is determined not by the total number of distinct terms but by the total number of relevant terms that cannot be eliminated.

Definition 3: Dominant rules are all the rules in the dominant model.

As shown in Section 2, Rules A1, and A2 are dominant rules for class 1, while Rules C1 and C2 are dominant rules for class 0.

Definition 4: A disjoint model is a model wherein there are no rules with overlapping condition spaces.

Models D and E are disjoint models. However, Model A is not a disjoint model, as it includes instances that satisfy both Rule A1 and Rule A2. Figure 1a shows that the respective condition spaces of Rule A1 and Rule A2 overlap with each other. On the other hand, Figure 1b shows that there is no overlapping space between Rule D1 and Rule D2.

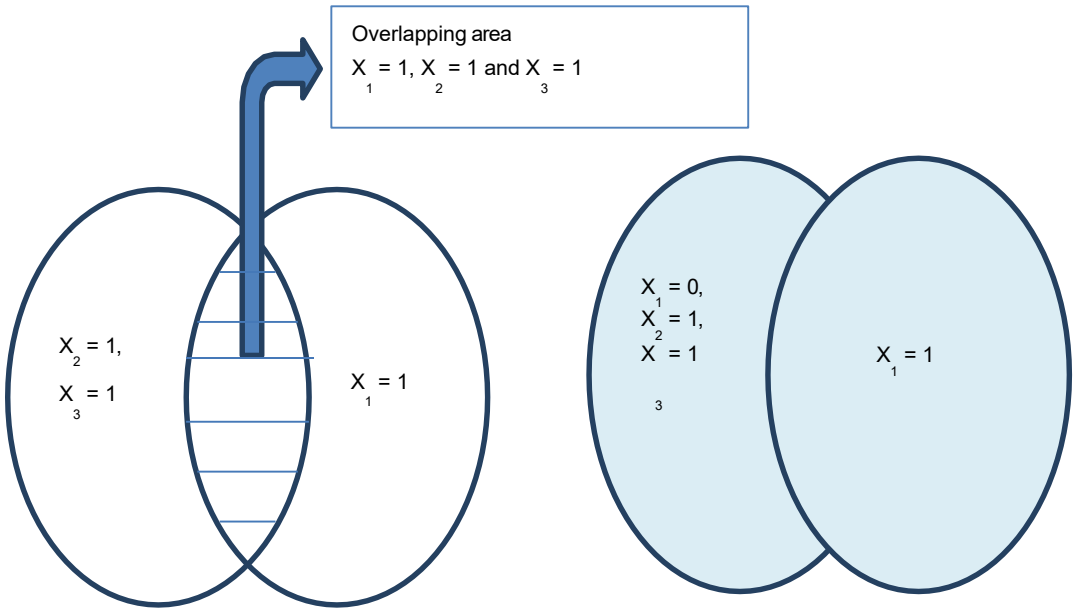
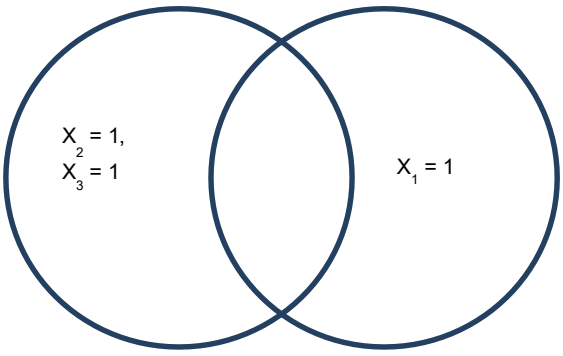


Figure 1a. Dominant Model Figure 1b. Disjoint Model

Definition 5: A child rule is a rule whose condition is the subset of another rule, i.e., (the parent rule).
Consider Model A and Model B in Section 2: Rule A1 is the parent rule of Rules B1, B2, and B3, whereas Rule A2 is the parent rule of Rule B4. In other words, Rules B1, B2, and B3 are the child rules of Rule A1, whereas Rule B4 is the child rule of Rule A2.
Definition 6: A complement model represents a class that differs from that of the given model.
If the given model is the model for Y = 1, then the complement model will be the model for Y = 0. For example, Model C in Section 2 is the complement model of Model A and vice versa. *Definition 7: An irrelevant rule is any rule that if eliminated would not change the response.*
Definition 8: An irrelevant term is any term in the condition of the rule that if eliminated would not change the response.
Consider Model F in Section 2: If we eliminate term X4 = 0 in Rule F1 and term X4 = 1 in Rule F2, the response would still be the same as it would had these terms not been eliminated. (An explanation is given in the following section.) Thus, both of these terms are irrelevant terms.
Next, we demonstrate how to convert any equivalent models into the dominant model.

Converting to the Dominant Model
The dominant model is unique and it is the model with the fewest condition terms that can be used to explain a dataset. Here, we propose the dominant model as a benchmark model for all the equivalent models. A step-by-step procedure for obtaining the dominant model from any equivalent models is given at the end of this subsection.

Terms from partitioned rule(s)
Suppose there are two rules: "if M_1 and M_2, then Y = 1" and "if M_1 and M^c, then Y = 1," where M^c is the complement term of M_2. The combination of these two rules is equivalent to the rule "if M_1, then Y = 1." As we can eliminate M_2 and M^c both are considered irrelevant terms. without changing the response,
As the example in Table 2 indicates, adding $X_4 = 0$ to Rule A1 generates Rule F1. Similarly, adding $X_4 = 1$ to Rule A1 generates Rule F2. Therefore, Model F is equivalent to Model A.



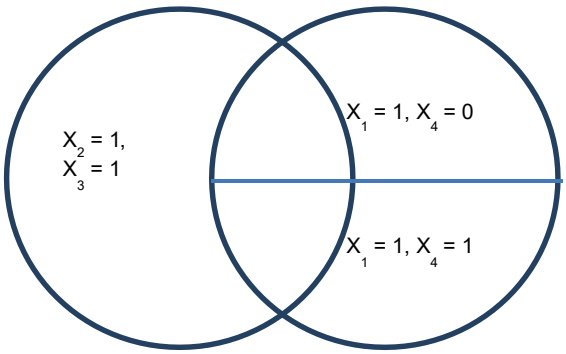


Figure 2a: Model A Figure 2b: Model F

As shown in Figure 2, one rule can be partitioned into multiple rules. It is, therefore, possible to generate numerous equivalent models as one rule can be partitioned in multiple ways.

Complement terms
Suppose the first rule is "if M_1, then Y = 1" and the second rule is "if M_2 and M_1^c, then Y = 1," where M_1^c is the

complement term of M_1. Note that the term M_1^c is irrelevant because it can be eliminated without changing the response. Consider the model condition of Rule D in Table 2: the complement A1 is $X_1 = 0$. This term in Rule D2 can be eliminated without changing the response of the model. Some examples of irrelevant terms related to this category are shown
in Figure 3.

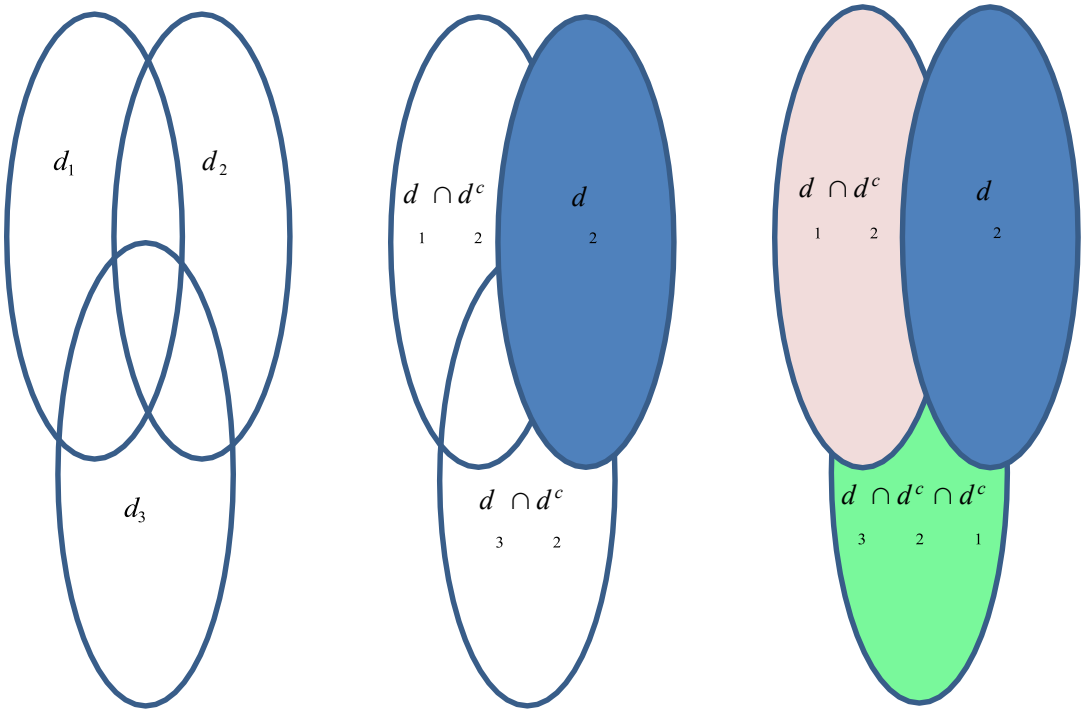


Figure 3a: Model with overlapping spaces

Figure 3b: Model wit overlapping spaces

Figure 3c: Model without overlapping spaces less

In Figure 3a, the model without any irrelevant terms is represented with three rules. The term d_2^c in Figure 3b and the terms d^c and d^c in Figure 3c are all irrelevant terms, as any of them can be eliminated without changing the response. Note that Figure 3c represents the disjoint model without any overlapping condition spaces.

The process stops when there are no longer any irrelevant terms in the model. The form after all the irrelevant rules and all the irrelevant terms have been deleted is the dominant model.

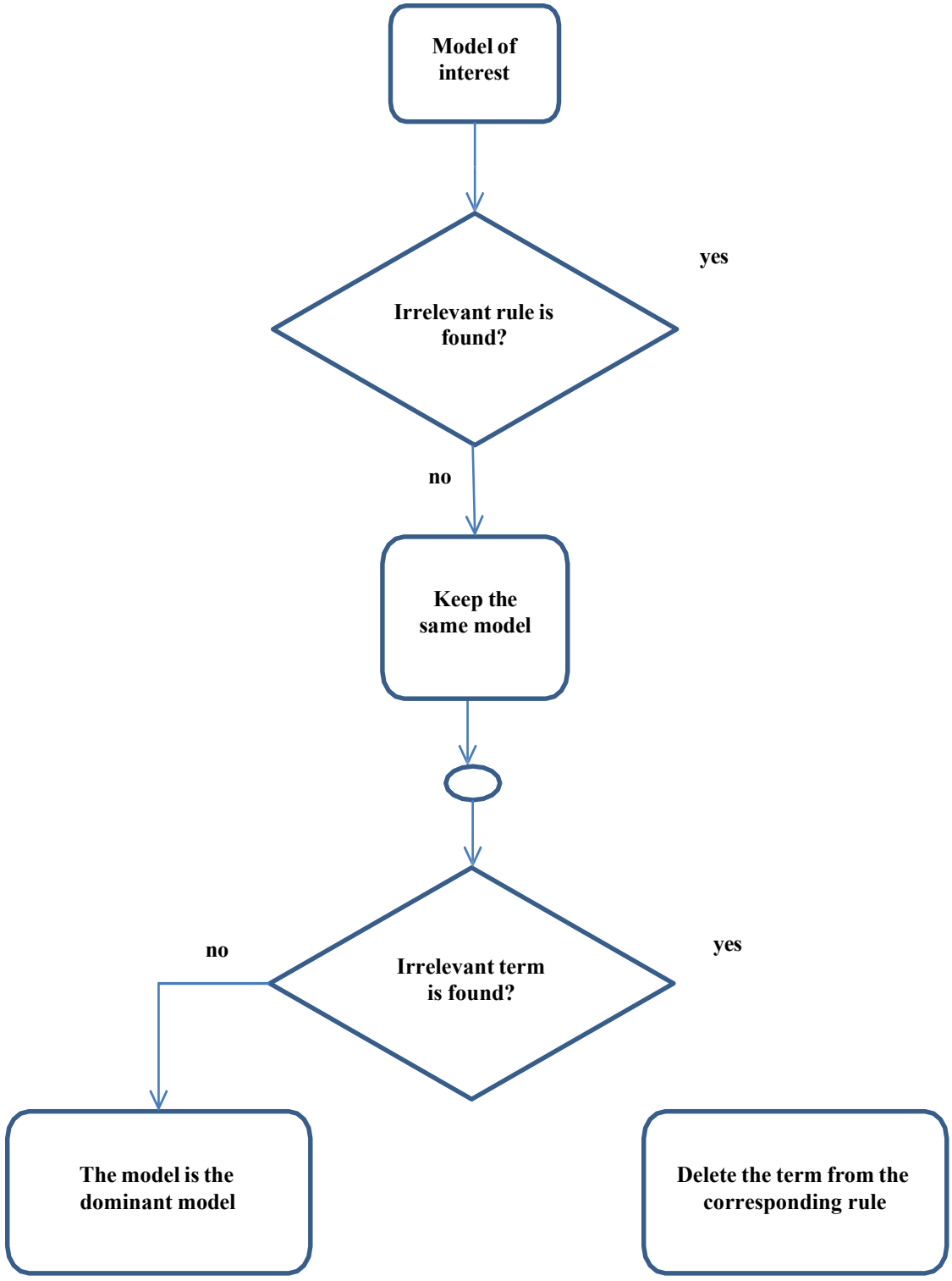


Figure 4: Decision chart for the search procedure for the dominant model Consequently, model equivalency is verified by comparing the dominant models identified

Corollary. All the condition terms in the dominant model appear in all the equivalent models.
The importance of uniqueness suggests that all the equivalent models can refer to this dominant model. The proof implies that all the equivalent models can be compressed into their simplest final form, i.e., the dominant model. We propose to use this model as a benchmark for explaining the relationships among variables for all the equivalent models. An example of how to convert any model into its dominant model follows.

Case Study: The MONK's dataset
The MONK's dataset was compiled by Sebastian Thrun of Carnegie Mellon University using propositional formulas over six factors (Thrun et al 1991). It can be obtained from the UCI machine learning website (http://archive.ics.uci.edu/ml/datasets.html). The MONK's problem is the basis for the first international comparison of learning algorithms.
This dataset has been used in numerous studies to compare methodologies. Among these, some of the papers that focus on finding the classification error rate without considering the model are Philip and Joseph (2000), Karaçali and Krim (2003), Shin (2006), Naval and Yusiong (2007), Zhang and Wang (2008), Appavu and Rajaram (2009), Qodmanan et al. (2011). Additionally, other studies search for models that represent the dataset (Etchells and Lisboa 2006; Guerreiro and Trigueiros 2010; Marghny and El-Semman 2005). These papers focus on the underlined rules that are given to the dataset without considering other possible equivalent models. According to these studies, methodologies that discover the given rules are competent. The problem is that the methodology can be competent even when the underlined rules are not found, as other equivalent models can be found instead. In the literature, many methodologies

have been studied in order to determine whether they can be used to find the given model (Etchells and Lisboa 2006; Guerreiro and Trigueiros 2010; Marghny and El-Semman 2005). We use the MONK's dataset to illustrate the effect of equivalent models with association rules. The derivation used to obtain the dominant model is also presented.
Background of the MONK's Dataset
The objective of the MONK's dataset is to classify 432 robots into two classes (class 0 and class 1) based on six attributes. The details of all the attributes are shown in Table 3. The given model assigns each robot to one of the two classes with a subset of the six attributes. Note that the given model for the MONK's dataset is designed on the basis of the following logic (the truth):
If the head shape and the body shape of the robot are the same, then the robot is in class 1, and
If the robot wears a red jacket, then the robot is in class 1.
We convert the original attributes into dummy variables, as shown in Table 2. The table also shows the role of equivalent rules in explaining the same dataset.

Table 2: Attributes for the MONK's Dataset

Attribute	Level	Binary Variable
head_shape	round, square, octagon	$X_1 = 1$ if head_shape is round, and $X_1 = 0$ otherwise. $X_2 = 1$ if head_shape is square, and $X_2 = 0$ otherwise.
body_shape	round, square, octagon	$X_3 = 1$ if body_shape is round, and $X_3 = 0$ otherwise. $X_4 = 1$ if body_shape is square, and $X_4 = 0$ otherwise.
is_smiling	yes, no	$X_5 = 1$ if smiling, and $X_5 = 0$ otherwise.
holding	sword, balloon, flag	$X_6 = 1$ if holding a sword, and $X_6 = 0$ otherwise. $X_7 = 1$ if holding a balloon, and $X_7 = 0$ otherwise.
jacket_color	red, yellow, green, blue	$X_8 = 1$ if jacket color is red, and $X_8 = 0$ otherwise. $X_9 = 1$ if jacket color is yellow, and $X_9 = 0$, otherwise. $X_{10} = 1$ if jacket color is green, and $X_{10} = 0$ otherwise.
has_tie	yes, no	$X_{11} = 1$ if wearing a tie, and $X_{11} = 0$ otherwise.

Equivalent Models for the MONK's dataset

If we follow the given logic with binary variables, the conditions are expressed in four rules—the dominant rules that form the dominant model, i.e., Model M, as shown in Table 3.

Table 3: Dominant Model of Class 1 for the MONK's Dataset (Model M)

Rule	Explanation
M1: If $X_1 = 1$ and $X_3 = 1$, then $Y = 1$.	If the robot has a round head shape and a round body shape, then the robot is in class 1.
M2: If $X_2 = 1$ and $X_4 = 1$, then $Y = 1$.	If the robot has a square head shape and a square body shape, then the robot is in class 1.
M3: If $X_1 = 0$, $X_2 = 0$, $X_3 = 0$, and $X_4 = 0$, then $Y = 1$.	If the robot has an octagonal head shape and an octagonal body shape, then the robot is in class 1.
M4: If $X_8 = 1$, then $Y = 1$.	If the robot wears a red jacket, then the robot is in class 1.

These four rules are the benchmark for comparing goodness of model and the criteria. Consider Model R, as shown in Table 4. There are seven rules, none of which appears in Model M (Table 3). Thus, by any existing criteria, Model R is not a good model. However, it can easily be verified that the responses generated by Model R are identical to those generated by Model M (Table 3). In other words, Model R is equivalent to Model M.

Table 4: Equivalent Model for the MONK's Dataset (Model R)

Rule	Explanation
R1: If $X_8 = 1$ and $X_1 = 1$, then $Y = 1$.	If the robot wears a red jacket and has a round head shape, then the robot is in class 1.
R2: If $X_8 = 0$, $X_1 = 1$, and $X_3 = 1$, then $Y = 1$.	If the robot has a round head shape, has a round body shape, and does not wear a red jacket, then the robot is in class 1.
R3: If $X_8 = 1$ and $X_4 = 1$, then $Y = 1$.	If the robot wears a red jacket and has a square body shape, then the robot is in class 1.
R4: If $X_8 = 0$, $X_2 = 1$, and $X_4 = 1$, then $Y = 1$.	If the robot has a square head shape, has a square body shape, and does not wear a red jacket, then the robot is in class 1.
R5: If $X_1 = 0$, $X_3 = 0$, and $X_8 = 1$, then $Y = 1$.	If the robot does not have a round head shape, does not have a round body shape, and wears a red jacket, then the robot is in class 1.
R6: If $X_1 = 0$, $X_2 = 0$, $X_3 = 0$, $X_4 = 0$, and $X_8 = 0$, then $Y = 1$.	If the robot has an octagonal head shape, has an octagonal body shape, and does not wear a red jacket, then the robot is in class 1.
R7: If $X_1 = 0$, $X_4 = 0$, and $X_8 = 1$, then $Y = 1$.	If the robot has a square head shape, has a square body shape, and wears a red jacket, then the robot is in class 1.

These two models are equivalent in that they possess identical responses. However, they differ in regard to how they explain the relationships between input variables and responses.

Dominant Model for MONK's Dataset

For Model R, the original seven rules cannot be deleted in the first step. Therefore, we consider irrelevant terms. The term that can be deleted from Rule R1 is X1 = 1, which leaves only the term X8 = 1 in Rule R1. Then Rule R3, Rule R5, and Rule R7 are deleted in the

consecutive steps, as they become the child rules of the modified Rule R1. Next, the term X8 = 0 is deleted from Rule R2, which leaves only the terms X1 = 1 and X3 = 1 in Rule R2. Again, as there is no irrelevant rule, we consider the next irrelevant term. The term X8 = 0 is deleted from Rule R4, which leaves only terms X2 = 1 and X4 = 1 in Rule R4. As there is no other irrelevant rule we consider the next irrelevant term. The term X8 = 0 is deleted from Rule R6, which leaves only the terms X1 = 0, X2 = 0, X3 = 0, and X4 = 0 in Rule R6. At this point, the dominant model, Model M, is obtained.

Note that the modified Rule R1 is identical to Rule M4, that the modified Rule R2 is identical to Rule M1, that the modified Rule R4 is identical to Rule M2, and that the modified Rule R6 is identical to Rule M3. As stated in Theorem 1, as this is the dominant model, it is unique. Note that we do not delete term X8 = 0 from Rule R2, Rule R4, and Rule R6 at the same time because deleting repeated terms does not always work with all equivalent models. In the next section, we show that the issue of equivalent model can happen to any real-life applications illustrated through the Lenses dataset (Cendrowska 1987).

Application: Lenses Dataset

The Lenses dataset was donated by Benoit Julien and it can be obtained from the UCI machine learning website (http://archive.ics.uci.edu/ml/datasets.html). The objective of the Lenses dataset is to classify 24 patients into two classes (class 0: does not need contact lenses; and class 1: needs contact lenses) based on four attributes. Note that the original data classify the patients into three classes (those who need hard contact lenses, those who need soft contact lenses, and those who do not need contact lenses). We modified this dataset for the purpose of illustration. The details of all the attributes are shown in Table 5.

Table 5: Attributes for the Lense Dataset

Attribute	Level	Binary Variable
Age of the patient	young, pre-presbyopic, presbyopic	$X_1 = 1$ if young, and $X_1 = 0$ otherwise. $X_2 = 1$ if pre-presbyopic, and $X_2 = 0$ otherwise.
Spectacle prescription	myope, hypermetrope	$X_3 = 1$ if myope, and $X_3 = 0$ otherwise.
Astigmatic	no, yes	$X_4 = 1$ if no, and $X_4 = 0$ otherwise.
Tear production rate	reduced, normal	$X_5 = 1$ if reduced, and $X_5 = 0$ otherwise.

To find a model for Lenses dataset, we use the CBA program (Liu et al. 1998) developed by the Department of Information Systems and Computer Sciences at the National University of Singapore (website: http://www.comp.nus.edu.sg/~dm2/). The model from the default setup by CBA is shown in Table 6.

Table 6: Model from association rules analysis for the Lenses Dataset

Rule	Explanation
If $X_1 = 1$ and $X_5 = 0$, then $Y = 1$.	If the patient is young and has a normal tear production rate, then the patient should wear contact lenses.
If $X_3 = 1$, $X_4 = 0$,and $X_5 = 0$, then $Y = 1$.	If the patient has a myope spectacle prescription, is astigmatic, and has a normal tear production rate, then the patient should wear contact lenses.
If $X_3 = 0$, $X_4 = 1$,and $X_5 = 0$, then $Y = 1$.	If the patient has a hypermetrope spectacle prescription, is not astigmatic, and has a normal tear production rate, then the patient should wear contact lenses.
If $X_3 = 1$, and $X_5 = 0$, then $Y = 1$.	If the patient has a myope spectacle prescription and has a normal tear production rate, then the patient should wear contact lenses.

We show via the Lenses dataset that association rules analysis produces an irrelevant rule and irrelevant term. The second rule and the term X3 = 0 in the third rule, as shown in Table 6, will be eliminated to obtain the dominant model, as shown in Table 7.

Table 7: Dominant Model of Class 1 for the Lenses Dataset

Rule	Explanation
If $X_1 = 1$ and $X_5 = 0$, then $Y = 1$.	If the patient is young and has a normal tear production rate, then the patient should wear contact lenses.
If $X_4 = 1$ and $X_5 = 0$, then $Y = 1$.	If the patient is not astigmatic and has a normal tear production rate, then the patient should wear contact lenses.
If $X_3 = 1$, and $X_5 = 0$, then $Y = 1$.	If the patient has a myope spectacle prescription and has a normal tear production rate, then the patient should wear contact lenses.

The illustration through the Lenses dataset shows the importance of obtaining the dominant model since the explanation from the dominant model in Table 7 is more sensible compared to the explanation from the model in Table 6.

Moreover, the phenomenon of equivalent models not only affects the association rules analysis, but it also affects the statistical techniques developed from the association rules analysis. For example, the model selection methodology for logistic regression by Changpetch and Lin (2013) performs variable selections among all the input variables and the variables generated from the selected association rules in order to build the optimal logistic regression model. Changpetch and Lin (2013) shows that the proposed methodology finds the dominant rules for the MONK's dataset. However, if the methodology finds the other equivalent models instead, the dominant model will be obtained by applying the procedure discussed in Section 3.

Parallel Effects to Decision Tree Analysis and the Statistical Model Selection Method

The effect of equivalent models is also relevant to decision tree analysis—a well-known method used to predict the class of response by decision rules. This technique has been used for many years with a wide range of applications (Yada et al. 2007; Wang and Liu 2008; Jaeger and Chen 2010; Lee 2010; Kim et al. 2011; Gepp et al. 2012; Delen et al. 2013).

Decision tree analysis often involves unavoidable irrelevant terms because of the hierarchical structure of this methodology. These irrelevant terms can appear in decision rules even with the most compact tree wherein the branches cannot be reduced any further. This study provides the methodology to establish the dominant model for decision tree analysis which is a different concept from that of finding the most compact decision tree (Quinlan 1986; Kienappel and Kneser 2001; Pal and Mather 2003; Pulkkinen and Koivisto 2008; Chandra and Varghese 2009; Cha and Tappert 2009). The present study shows that we can eliminate irrelevant terms or rules after obtaining the decision rules from the decision tree. Even though we obtain the most compact tree, the irrelevant rules and terms are tentative because of the hierarchical structure. To illustrate, we present examples of relevant terms and how to obtain the dominant models through the Lymphography dataset (Polat and Güneş 2007).

Application: Lymphography Dataset

In this section, we use the Lymphography dataset from the University Medical Centre at the Institute of Oncology in Ljubljana, Yugoslavia, which was graciously provided by M. Zwitter and M. Soklic. Here, we illustrate an application of our proposed method by eliminating the irrelevant terms from the decision rules.

The Lymphography dataset comprises records of lymph disease diagnoses in four classes: normal, metastases, malign lymph, and fibrosis. The dataset consists of 148 total instances: 2 normal diagnoses, 81 metastases diagnoses, 61 malign lymph diagnoses, and 4 fibrosis diagnoses. As there were so few normal and fibrosis diagnoses, we did not include them in our analysis. The modified dataset, therefore, comprised two classes and 142 instances. The goal was to determine the response classes for metastases and malign lymph. There were 18 numerically valued attributes, which we converted into binary variables, listed in Table 8.

Table 8: Responses and Attributes for the Lymphography Dataset

Attribute	Levels	Binary Variables
Diagnosis class	malign lymph, metastases	Y = 1 if the class is malign lymph Y = 0 if the class is metastases
Lymphatic	arched, deformed, displaced	X1 = 1 if arched, and X1 = 0 otherwise X2 = 1 if deformed, and X2 = 0 otherwise
Block of afferent	yes, no	X3 = 1 if yes, and X3 = 0 otherwise
Block of lymph c	yes, no	X4 = 1 if yes, and X4 = 0 otherwise
Block of lymph s	yes, no	X5 = 1 if yes, and X5 = 0 otherwise
Bypass	yes, no	X6 = 1 if yes, and X6 = 0 otherwise
Extravagates	yes, no	X7 = 1 if yes, and X7 = 0 otherwise
Regeneration	yes, no	X8 = 1 if yes, and X8 = 0 otherwise
Early uptake	yes, no	X9 = 1 if yes, and X9 = 0 otherwise
Lymph node dimensions	1, 2	X10 = 1 if 2, and X10 = 0 otherwise
Lymph nodes enlarged	1, 2, 3, 4	X11 = 1 if 1, and X11 = 0 otherwise X12 = 1 if 2, and X12 = 0 otherwise X13 = 1 if 3, and X13 = 0 otherwise
Changes in lymph	bean, oval, round	X14 = 1 if bean, and X14 = 0 otherwise X15 = 1 if oval, and X15 = 0 otherwise
Defects in node	no, lacunars, lacunars marginal, lacunars central	X16 = 1 if no, and X16 = 0 otherwise X17 = 1 if lacunars, and X17 = 0 otherwise X18 = 1 if lacunars marginal, and X18 = 0 otherwise

Changes in node	no, lacunars, lacunars marginal, lacunars central	X19 = 1 if no, and X19 = 0 otherwise X20 = 1 if lacunars, and X20 = 0 otherwise X21 = 1 if lacunars marginal, and X21 = 0 otherwise
Changes in structure	no, grainy, drop-like, coarse, diluted, reticular, stripped, faint	–
Special forms	no, chalices, vesicles	X22 = 1 if no, and X22 = 0 otherwise X23 = 1 if chalices, and X23 = 0 otherwise
Dislocation of node	yes, no	X24 = 1 if yes, and X24 = 0 otherwise
Exclusion of node	yes, no	X25 = 1 if yes, and X25 = 0 otherwise
Number of nodes	0–9, 10–19, 20–29, 30–39, 40–49, 50–59, 60–69, > = 70	X26 = 1 if 0–9, and X26 = 0 otherwise X27 = 1 if 10–19, and X27 = 0 otherwise

In the original dataset, there are more levels within some variables. However, when we deleted the six instances of normal and fibrosis diagnoses, some levels were eliminated. We decided to eliminate the attribute change in structure, as we did not have a solid criterion on which to discretize it because it had eight levels within the binary variable. We discretized number of nodes by dividing the observations into three groups based on distribution characteristics.

Figure 5 shows one of the decision trees for the Lymphography dataset. With this decision tree, irrelevant terms are attached to the decision rules, as shown in Table 9.

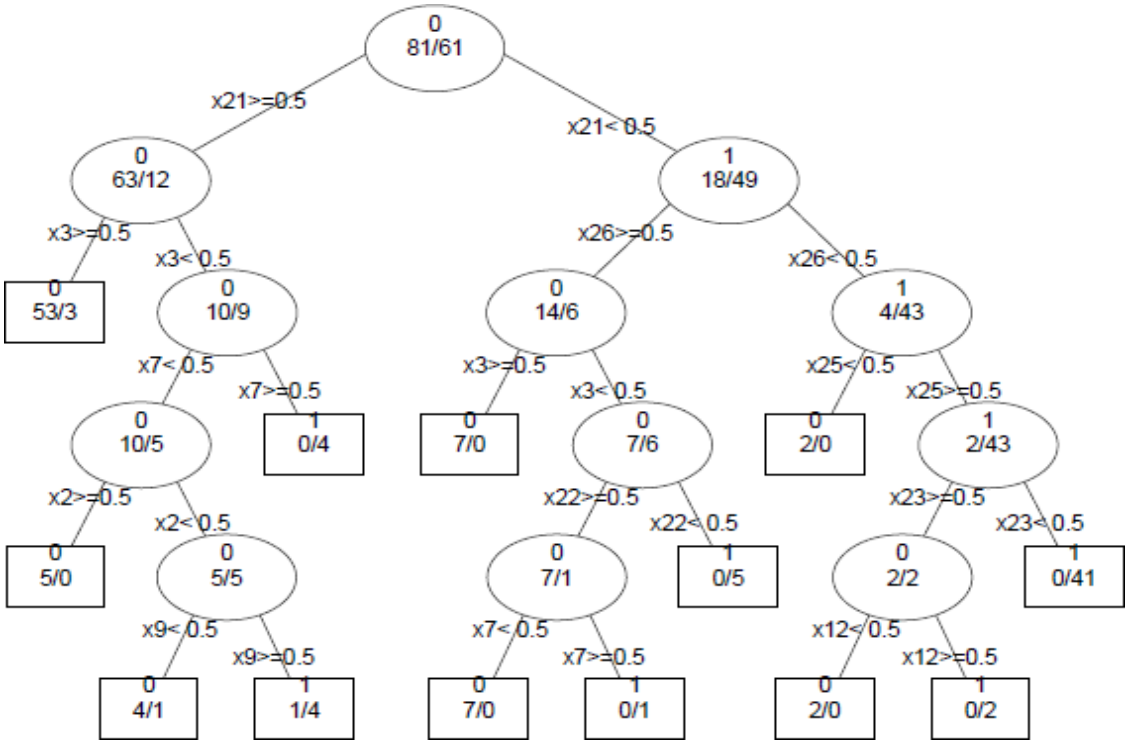


Figure 5: Decision tree for the Lymphography dataset

We show via the Lymphography dataset that decision tree analysis produces irrelevant terms. As shown in Table 9, the term X22 = 1 in the first rule, the term X23 = 1 in the third rule, and the term X7 = 0 in the sixth rule are eliminated to obtain the dominant model, as shown in Table 10.

Table 9: Model from Decision Tree Analysis for the Lymphography Dataset

Rule	Explanation
If $X_{21} = 0$, $X_{26} = 1$, $X_3 = 0$, $X_{22} = 1$, and $X_7 = 1$, then Y = 1.	If changes in node is not lacunars marginal, number of nodes is 0–9, block of afferent is no, special forms is no, and extravagates is yes, then the patient is in the malign lymph class.
If $X_{21} = 0$, $X_{26} = 1$, $X_3 = 0$, and $X_{22} = 0$, then Y = 1.	If changes in node is not lacunars marginal, number of nodes is 0–9, block of afferent is no,
	and special forms is either chalices or vesicles, then the patient is in the malign lymph class.
If $X_{21} = 0$, $X_{26} = 0$, $X_{25} = 1$, $X_{23} = 1$, and $X_{12} = 1$, then Y = 1.	If changes in node is not lacunars marginal, number of nodes is not 0–9, exclusion of node is yes, special forms is chalices, and lymph nodes enlarged is 2, then the patient is in the malign lymph class.
If $X_{21} = 0$, $X_{26} = 0$, $X_{25} = 1$, and $X_{23} = 0$, then Y = 1.	If changes in node is not lacunars marginal, number of nodes is not 0–9, exclusion of node is yes, and special forms is not chalices, then the patient is in the malign lymph class.
If $X_{21} = 1$, $X_3 = 0$, and $X_7 = 1$, then Y = 1.	If changes in node is lacunars marginal, block of afferent is no, and extravagates is yes, then the patient is in the malign lymph class.
If $X_{21} = 1$, $X_3 = 0$, $X_7 = 0$, $X_2 = 0$, and $X_9 = 1$, then Y = 1.	If changes in node is lacunars marginal, block of afferent is no, extravagates is no, lymphatic is not deformed, and early uptake is yes, then the patient is in the malign lymph class.

Table 10: Dominant Model of Class 1 for the Lymphography Dataset

Rule	Explanation
If $X_{21} = 0$, $X_{26} = 1$, $X_3 = 0$, and $X_7 = 1$, then Y = 1.	If changes in node is not lacunars marginal, number of nodes is 0–9, block of afferent is no, and extravagates is yes, then the patient is in the malign lymph class.
If $X_{21} = 0$, $X_{26} = 1$, $X_3 = 0$, and $X_{22} = 0$, then Y = 1.	If changes in node is not lacunars marginal, number of nodes is 0–9, block of afferent is no, special forms is either chalices or vesicles, then the patient is in the malign lymph class.
If $X_{21} = 0$, $X_{26} = 0$, $X_{25} = 1$, and $X_{12} = 1$, then Y = 1.	If changes in node is not lacunars marginal, number of nodes is not 0–9, exclusion of node is yes, and lymph nodes enlarged is 2, then the patient is in the malign lymph class.
If $X_{21} = 0$, $X_{26} = 0$, $X_{25} = 1$, and $X_{23} = 0$, then Y = 1.	If changes in node is not lacunars marginal, number of nodes is not 0–9, exclusion of node is yes, and special forms is not chalices, then the patient is in the malign lymph class.
If $X_{21} = 1$, $X_3 = 0$, and $X_7 = 1$, then Y = 1.	If changes in node is lacunars marginal, block of afferent is no, and extravagates is yes, then the patient is in the malign lymph class.
If $X_{21} = 1$, $X_3 = 0$, $X_2 = 0$, and $X_9 = 1$, then Y = 1.	If changes in node is lacunars marginal, block of afferent is no, lymphatic is not deformed, and early uptake is yes, then the patient is in the malign lymph class.

The application demonstrates that decision trees always generate irrelevant terms, which can be eliminated by our method. Accordingly, the dominant model can be found for any model from a decision tree.

As a consequence, the statistical techniques that use the variables generated from decision tree analysis will have this effect as well. For example, Kim (2009) combined the logistic regression model with decision tree analysis. Each branch of the decision tree becomes the variable (or cluster) for logistic regression model selection. Our approach can help to eliminate the irrelevant ones to formulate a consistent and concise explanation of the dataset.

Conclusion

The literature has not sufficiently addressed the idea of model equivalence, which is predicated on two or more models producing equal results. Its impacts may be seen in decision tree analysis and other similar statistical methods in addition to association rules analysis. According to our investigation, two models may accurately reflect the same dataset even if they don't seem to be related. Consequently, as was previously said, neglecting to take equivalence into consideration might result in outcomes that are deceptive when it comes to creating a technique and drawing comparisons.

However, the result of using distinct sets of rules to describe the same dataset is that we have many explanations

for the dataset, some of which may not be consistent. Our main discovery is that, despite the fact that several models may reflect the same dataset, there is only one distinct model—the "dominant" model—to which all comparable models may be referred. The advantage of the dominant model is that it may be used as a benchmark to explain the connections between the variables and to confirm model equivalence, eliminating unfair comparisons between the criteria and rules selection procedures.

The phenomena of comparable models was investigated in this study. First, we demonstrated that if this phenomena is ignored, false findings are likely to be produced. The notion of the dominating model among all comparable models was then examined and defined. We also explained how we arrived at the dominating model. Third, in order to describe the connections between variables, we suggested utilizing the dominant model as the benchmark model. We anticipate that this finding will now be taken into account in decision tree analysis, association rules analysis, and other relevant statistical and data analysis techniques.

Although we used prediction rules (supervised learning) to illustrate the model equivalency phenomenon, they are often class association rules (Li et al. 2001). We believe that more generic association models probably take model equivalency into account.

References

Agrawal, R. and Srikant, S. (1994). Fast algorithms for mining association rules. Proceedings of the ACM SIGMOD Conference on Management of Data, 487–499.

Appavu, S. and Rajaram, R. (2009). Knowledge-based system for text classification using ID6NB algorithm. Knowledge-Based Systems 22: 1–7.

Appice, A., Ceci, M., Lanza, A., Lisi, F. A., and Malerba, D. (2003). Discovery of spatial association rules in georeferenced census data: A relational mining approach. Intelligent Data Analysis 7(6): 541–566.

Bastide, Y., Pasquier, N., Taouil, R., Stumme, G., and Lakhal, L. (2000). Mining minimal non-redundant association rules using frequent closed itemsets. Lecture Notes in Computer Science 1861: 972–986.

Bayardo Jr, R. J. (1997). Brute-force mining of high-confidence classification rules. Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, Newport Beach, CA, 123–126.

Becquet, C., Blachon, S., Jeudy, B., Boulicaut, J. F., and Gandrillon, O. (2002). Strong association rule mining for large gene expression data analysis: A case study on human SAGE data. Genome Biology 3(12): 1-16.

Cendrowska, J. (1987). PRISM: An algorithm for inducing modular rules. International Journal of Man-Machine Studies 27: 349-370.

Cha, S. and Tappert, C. (2009). A generic algorithm for constructing compact binary decision trees. Journal of Pattern Recognition Research 1: 1–13.

Chandra, B. and Varghese, P. P. (2009). Moving towards efficient decision tree construction. Information Sciences 179: 1059–1069.

Changpetch, P. and Lin D. K. J. (2013). Model selection for logistic regression via association rules analysis. Journal of Computational Statistics and Simulations 83(8): 1415– 1428.

Delen, D., Kuzey, C. and Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. Expert Systems with Applications 40: 3970–3983.

Dong, G., Zhang, X., Wong, L. and Li, J. (1999). CAEP: Classification by aggregating emerging patterns. Proceedings of the Second International Conference on Discovery Science (DS'99), Japan, 30–42.

Dopfer, K. and Potts, J. (2004). Evolutionary realism: A new ontology for economics. Journal of Economic Methodology 11: 195–212.

Etchells, T. A. and Lisboa, P. J. G. (2006). Orthogonal Search-based Rule Extraction (OSRE) method for trained neural networks: A practical and efficient approach. IEEE Transactions on Neural Networks 17(2): 374–384.

Garcia, P., Amandi, A., Schiaffino, S. and Campo, M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. Computer and Education Journal 49: 794– 808.

Gepp, A., Wilson, J. H., Kumar, K. and Bhattacharya, S. (2012). A Comparative Analysis of Decision Trees Vis-à-vis Other Computational Data Mining Techniques in Automotive Insurance Fraud Detection. Journal of Data Science 10(3): 537-561.

Grudzinski, K., Grochowski, M. and Duch, W. (2010). Pruning classification rules with reference vector selection methods. LNCS 6113: 347–354.
Guerreiro, J. and Trigueiros, D. (2010). A unified approach to the extraction of rules from artificial neural networks and support vector machines. Advanced Data Mining and Applications, 34–42.
Jaeger, S. and Chen, S. (2010). Information fusion for biological prediction. Journal of Data Science 8(2): 269-288.
Karaçali, B. and Krim, H. (2003). Fast minimization of structural risk by nearest neighbor rule. IEEE Transactions on Neural Network 14(1), 127–137.
Kienappel, A. K. and Kneser, R. (2001). Designing very compact decision trees for grapheme-to-phoneme transcription. In Proceedings Eurospeech (Scandinavia), 1911–1914.
Kim, M. (2009). Two-stage logistic regression model. Expert Systems with Applications 36: 6727–6734.
Kim, S. S., Timothy, D. J. and Hwang, J. (2011). Understanding Japanese tourists' shopping preferences using the decision tree analysis method. Tourism management 32: 544–554.
Kuo, R. J., Lin, S. Y. and Shih, C. W. (2007). Mining association rules through integration of clustering analysis and ant colony system for health insurance database in Taiwan. Expert Systems with Applications 33(3): 794–808.
Lee, A. J. T., Hong, R. W., Ko, W. M., Tsao, W. K. and Lin, H. H. (2007). Mining spatial association rules in image databases. Information Sciences 177(7): 1593–1608.
Lee, S. (2010). Using data envelopment analysis and decision trees for efficiency analysis and recommendation of B2C controls. Decision Support Systems 49: 486–497.
Li, J. (2006). On optimal rule discovery. IEEE Transactions on Knowledge and Data Engineering 18(4): 460–471.
Li, W., Han, J. and Pei, J. (2001). CMAR: Accurate and efficient classification based on multiple class-association rules. Proceedings of the 2001 International Conference on Data Mining (ICDM'01), San Jose, CA, 369–376.
Liu, B., Hsu, W. and Ma, Y. (1998). Integrating classification and association rule mining. Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD'98), New York, 80–86.
Liu, K. H., Weng, M. F., Tseng, C. Y., Chuang, Y. Y., Chen, M. S. (2008). Association and temporal rule mining for post-processing of semantic concept detection in video. IEEE TMM 10(2): 240–251.
Marghny, M. H. and El-Semman, I. E. (2005). Extracting logical classification rules with gene expression programming: Microarray case study. Proceedings of the International Conference on Artificial Intelligence and Machine Learning (AIML'05), Cairo, Egypt, 11– 16.
Matsumoto, K. (1998). An experimental agricultural data mining system. Proceedings of the First International Conference on Discovery Science (DS'98), Japan, 439–440.
McIver, D. K. and Friedl, M. A, (2002). Using prior probabilities in decision tree classification of remotely sensed data. Remote Sensing of Environment 81: 253–261.
Merceron, A. and Yacef, K. (2008). Interestingness measures for association rules in educational data. Proceedings of Educational Data Mining Conference (EDM'08), 57–66.
Mossong, J., Hens, N., Jit, M. et al. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Medicine 5(3): 381–391.
Naval Jr, P. C. and Yusiong, J. P. T. (2007). An evolutionary multi-objective neural network optimizer with bias-based pruning heuristic. Advances in Neural Networks, 174–183.
Pal, M. and Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sensing of Environment 86: 554–565.
Philip, N. S. and Joseph, K. B. (2000). Boosting the differences: A fast Bayesian classifier neural network. Intelligent Data Analysis 4: 463–473.
Polat, K. and Güneş, S. (2007). Automatic determination of diseases related to lymph system from lymphography data using principles component analysis (PCA), fuzzy weighting pre- processing and ANFIS. Expert Systems with Applications 33: 636–641.
Pulkkinen, P. and Koivisto, H. (2008). Fuzzy classifier identification using decision tree and multiobjective evolutionary algorithms. International Journal of Approximate Reasoning 48: 526–543.
Qodmanan, H. R., Nasiri, M. and Minaei-Bidgoli, B. (2011). Multi objective association rule mining with genetic algorithm without specifying minimum support and minimum confidence. Expert Systems with Applications