Using Extreme Prior Probabilities on the Naive Credal Classifier

Serafín Moral-García[a,*], Javier G. Castellano[a], Carlos J. Mantas[a], Joaquín Abellán[a]

[a]*Department of Computer Science and Artificial Intelligence University of Granada, Granada, Spain*

Abstract

The Naive Credal Classifier (NCC) was the first method proposed for Imprecise Classification. It starts from the known Naive Bayes algorithm (NB), which assumes that the attributes are independent given the class variable. Despite this unrealistic assumption, NB and NCC have been successfully used in practical applications. In this work, we propose a new version of NCC, called Extreme Prior Naive Credal Classifier (EP-NCC). Unlike NCC, it takes into consideration the lower and upper prior probabilities of the class variable in the estimation of the lower and upper conditional probabilities. We demonstrate that, with our proposed EP-NCC, the predictions are more informative than with NCC without increasing the risk of making erroneous predictions. An experimental analysis carried out in this work shows that EP-NCC significantly outperforms NCC and obtains statistically equivalent results to the algorithm proposed so far for Imprecise Classification based on decision trees, even though EP-NCC is computationally simpler. Therefore, EP-NCC is more suitable to be applied to large datasets for Imprecise Classification than the methods proposed so far in this field. This is an important issue in favor of our proposal due to the increasing amount of data in every area.

Keywords:
Imprecise Classification, Naive Credal Classifier, Extreme Prior Naive Credal Classifier, lower and upper prior probabilies, non-dominated states set

*Corresponding Author

Email addresses: seramoral@decsai.ugr.es (Serafín Moral-García), fjgc@decsai.ugr.es (Javier G. Castellano), cmantas@decsai.ugr.es (Carlos J. Mantas), jabellan@decsai.ugr.es (Joaquín Abellán)

1. Introduction

Classification [1] is an important task within Machine learning that aims to predict, for an instance described via a set of *attributes* or *features*, the value of a variable under study known as the *class variable*. The algorithms that make such predictions are usually called *classifiers*. The prediction made by a classifier often consists of a single class value. Nonetheless, in many situations, it is more appropriate that classifiers predict a set of states of the class variable because the available information is not reliable enough to point at only one class value. This is known as an *imprecise prediction*. In *decision-making*, this might imply fewer erroneous decisions that lead to negative consequences. Classifiers that make imprecise predictions are called *imprecise classifiers* [2].

When an imprecise classifier is used, it might obtain a set of states of the class variable. This set is composed of those values of the class variable not "defeated" by the rest according to a criterion, which is known as the *dominance criterion*. The predicted set of class values is called the set of *non-dominated states*. Intuitively, an evaluation measure of an imprecise classifier must take into consideration whether the real class value belongs to the set of non-dominated states and how informative is this set of non-dominated states, which is measured by its cardinality.

For building an imprecise classifier, it is more suitable to use models based on imprecise probabilities than the ones that use classical Probability Theory. Specifically, the Imprecise Dirichlet Model (IDM) [3], belonging to the reachable probability intervals theory [4], has been widely used in imprecise classifiers [5, 6, 7, 8].

So far, few algorithms for Imprecise Classification have been proposed. The first one of them was the *Naive Credal Classifier* (NCC) [2, 9], which starts from the Naive Bayes algorithm (NB) [10] for standard classification. NCC employs models based on imprecise probabilities combined with the naive assumption [10], which establishes that all the features are independent given the class variable, to provide imprecise predictions. NCC can make reliable predictions with fewer data and, thus, it is appropriate to handle uncertainty with few data [2, 9].

Obviously, in many cases, the naive assumption is not satisfied. Despite this, in precise classification, the NB algorithm has provided results comparable with other more sophisticated methods, especially when the correlation among the attributes is not very high [11, 12, 13]. Indeed, the NB algo-

rithm has been effectively used in practice [14, 15, 16]. Also, for Imprecise Classification, NCC has been successfully employed in practical applications [17, 18, 19]. Moreover, it is convenient to remark that NCC, due to its naive assumption, is much faster than other algorithms for Imprecise Classification and requires far less computational cost.

NCC is based on the Bayes theorem. In NB classifiers, this formula is employed with the naive assumption. The estimation of the conditional probabilities of the attributes given the class variable is crucial in the evaluation of the Bayes formula [20, 21]. In [20], the performance of the NB algorithm for precise classification is improved by considering the prior probability of the class variable in the estimation of the conditional probabilities.

In this research, we propose a new version of the NCC algorithm called Extreme Prior Naive Credal Classifier (EP-NCC). This new method, unlike NCC, for the estimation of the lower and upper conditional probabilities of the class variable, takes the lower and upper prior probabilities into account. We show that, with our proposed algorithm, the non-dominated states set is often smaller than with the original NCC. In this way, it solves the drawback found in [5] about the NCC method: it is not very informative, i.e it predicts too many class values. Since EP-NCC predicts fewer class values than NCC, the risk of making erroneous predictions might be higher with the first method. Nevertheless, EP-NCC takes the extreme prior probabilities of the class variable into account to reduce the number of predicted states and, thus, the risk of incorrect predictions is not much higher than with NCC.

We carry out exhaustive experimentation to compare the performance of NCC, EP-NCC, and Imprecise Credal Decision Tree (ICDT), an algorithm for Imprecise Classification based on decision trees proposed by Abellán and Masegosa in [5].[1] This experimental analysis shows that our proposed algorithm EP-NCC outperforms the already existing NCC and that EP-NCC obtains statistically equivalent results to the ones achieved by ICDT. More specifically: EP-NCC obtains significantly better results than NCC in the metrics corresponding to the number of predicted states; the performance of both algorithms is equivalent in the metrics associated with making correct predictions; even though EP-NCC is not as accurate as ICDT, the former method is more informative than the latter one as it predicts fewer states of

[1] This method is not as simple as NCC and EP-NCC because it does not use the naive assumption, and it is based on a recursive building procedure based on decision trees.

the class variable. The experimentation also shows that the processing time for our proposed EP-NCC algorithm is considerably lower than for ICDT.

This paper is arranged as follows: Section 2 describes the Imprecise Dirichlet Model, the dominance criteria used on this model, and the Naive Credal Classifier. The proposed Extreme Prior Naive Credal Classifier is exposed in Section 3. Section 4 details the experimental study carried out in this work. Conclusions are given in Section 5.

2. Background

Within this section, we use the following notation:
Let X be a discrete variable whose set of possible values is $\{x_1, \ldots, x_t\}$:

- Let $P(X = x_i)$ denote the probability that X takes the x_i value, $\forall i = 1, 2, \ldots, t$, being P a probability measure.
- If p is a probability distribution on X, $p(x_i)$ indicates the value that p takes on x_i, $\forall i = 1, 2, \ldots, t$.
- Let $\mathcal{P}(X)$ denote a closed and convex set of probability distributions on X, also called credal set.
- Let Y be another discrete variable. For each $i = 1, 2, \ldots, t$, let $\mathcal{P}(Y \mid x_i)$ denote a credal set on Y conditioned to X takes the x_i value, $\forall i = 1, 2, \ldots, t$.

2.1. Imprecise Dirichlet Model

2.1.1. Probability Intervals Theory:

Suppose that X is a discrete variable that takes values in $\{x_1, x_2, \ldots, x_t\}$. Probability Intervals Theory [4] establishes that the probability that X takes each possible value x_i belongs to a given interval $[l_i, u_i]$, $\forall i = 1, 2, \ldots t$. The set of probability intervals $\mathcal{I} = \{[l_i, u_i], \quad i = 1, 2, \ldots, t\}$ gives rise to the following credal set on X [4]:

$$\mathcal{P}(\mathcal{I}) = \{p \mid l_i \leq p(\{x_i\}) \leq u_i, \quad \forall i = 1, 2, \ldots, t\}. \tag{1}$$

In order to avoid that the intervals are too wide, the concept of *reachability* is introduced in [4]:

Definition 1. *"A set of probability intervals on X $\mathcal{I} = \{[l_i, u_i], \quad i = 1, 2, \ldots, t\}$ is said to be reachable if, $\forall i \in \{1, 2, \ldots, t\}$ and for each $v_i \in \{l_i, u_i\}$, there exists $p \in \mathcal{P}(I)$ such that $p(\{x_i\}) = v_i$."*

The following result allows checking the reachability of a set of probability intervals [4]:

Proposition 1. *$\mathcal{I}$ is reachable if, and only if, the following two conditions are verified:*

$$u_i + \sum_{j=1, j\neq i}^{t} l_j \leq 1, \quad \forall i = 1, 2, \ldots, t,$$

$$l_i + \sum_{j=1, j\neq i}^{t} u_j \geq 1, \quad \forall i = 1, 2, \ldots, t.$$

2.1.2. Probability intervals from the Imprecise Dirichlet Model:

Let X be a discrete attribute whose set of possible values is $\{x_1, x_2, \ldots, x_t\}$. Suppose that we have a sample of N independent and identically distributed outcomes of X.

The Imprecise Dirichlet Model (IDM) [3] estimates that:

$$P(X = x_i) \in I_i = \left[\frac{n_i}{N+s}, \frac{n_i + s}{N+s}\right], \tag{2}$$

where n_i is the number of cases in the sample for which $X = x_i$, $\forall i = 1, 2, \ldots, t$, and $s > 0$ is a given hyperparameter of the model.

The set of probability intervals $\mathcal{I} = \left\{I_i = \left[\frac{n_i}{N+s}, \frac{n_i}{N+s}\right], \quad i = 1, 2, \ldots, t\right\}$ is reachable [22]. It gives rise to the following credal set on X [22]:

$$\mathcal{P}_{IDM}(X) = \left\{p \mid \frac{n_i}{N+s} \leq p(\{x_i\}) \leq \frac{n_i + s}{N+s}, \quad \forall i = 1, 2, \ldots, t\right\}. \tag{3}$$

An important issue is the choice of the s hyperparameter. The intervals are wider as s is higher. This hyperparameter determines the speed of convergence of lower and upper probabilities for each x_i as the sample size is larger. Two values are suggested in [3]: $s = 1$ and $s = 2$.

2.2. Dominance criteria on the IDM

In Imprecise Classification, we employ a *dominance criterion* to determine the non-dominated states set, i.e the set of values of the class variable that are not "defeated" under probabilistic terms by another. For this purpose, under the IDM, the interval bounds can be used.

Let c_i and c_j be two possible values of the class variable. With the IDM, the two following criteria are frequently used:

- Suppose that $[l_i, u_i]$ and $[l_j, u_j]$ are the probability intervals corresponding to c_i and c_j, respectively. It is said that there is *stochastic dominance* of c_i on c_j if $l_i \geq u_j$.
- When the information about the class variable C is expressed via a non-empty credal set $\mathcal{P}$, it is said that there exists *credal dominance* of c_i on c_j if $p(\{c_i\}) \geq p(\{c_j\})$, $\forall p \in \mathcal{P}$.

Stochastic dominance always implies credal dominance, but the inverse is not true [23]. Both dominance criteria are equivalent when they are applied on reachable probability intervals [23]. Thereby, since IDM probability intervals are reachable, with this model, credal and stochastic dominance criteria always provide the same results.

2.3. The Naive Credal classifier

Let C be the class variable and $\{c_1, c_2, \ldots, c_K\}$ its set of possible values. Suppose that we have n attributes $X^1, X^2, \ldots, X^n$, and that each attribute X^i takes values on a finite set $\left\{x_1^i, x_2^i, \ldots, x_{t_i}^i\right\}$, $\quad \forall i = 1, 2, \ldots, n$.

The Naive Credal Classifier (NCC) [2] is based on the naive assumption [10], according to which the attributes are independent given the class variable:

$$\begin{aligned} P\left(C = c_j \mid X^1 = x_{r_1}^1, X^2 = x_{r_2}^2, \ldots, X^n = x_{r_n}^n\right) = \\ \frac{P\left(C = c_j, X^1 = x_{r_1}^1, X^2 = x_{r_2}^2, \ldots, X^n = x_{r_n}^n\right)}{P\left(X^1 = x_{r_1}^1, X^2 = x_{r_2}^2, \ldots, X^n = x_{r_n}^n\right)} \sim \\ P\left(C = c_j, X^1 = x_{r_1}^1, X^2 = x_{r_2}^2, \ldots, X^n = x_{r_n}^n\right) = \\ P\left(C = c_j\right) P\left(X^1 = x_{r_1}^1, X^2 = x_{r_2}^2, \ldots, X^n = x_{r_n}^n \mid C = c_j\right) = \\ P\left(C = c_j\right) \prod_{i=1}^{n} P\left(X^i = x_{r_i}^i \mid C = c_j\right), \\ \forall j = 1, \ldots, K, \quad \forall r_i \in \{1, 2 \ldots, t_i\}, \quad \forall i = 1, 2, \ldots, n. \end{aligned} \tag{4}$$

The proportionality relation indicated via the symbol $\sim$ is because

$$\arg\max_{1\leq j\leq K}\left(\frac{P\left(C=c_j,X^1=x^1_{r_1},X^2=x^2_{r_2},\ldots,X^n=x^n_{r_n}\right)}{P\left(X^1=x^1_{r_1},X^2=x^2_{r_2},\ldots,X^n=x^n_{r_n}\right)}\right)=$$
$$\arg\max_{1\leq j\leq K}\left(P\left(C=c_j,X^1=x^1_{r_1},X^2=x^2_{r_2},\ldots,X^n=x^n_{r_n}\right)\right).$$

Definition 2. *Let $\mathcal{P}(C)$ be a credal set on the class variable C. For each $i=1,2,\ldots,n$, $k=1,2,\ldots,K$, let $\mathcal{P}(X^i\mid c_k)$ be a credal set on the X^i attribute conditioned to $C=c_k$. These credal sets are called local credal sets [2].*

The NCC algorithm considers, for each $k=1,2,\ldots,K$, the set of joint probability distributions $\mathcal{P}(c_k,X^1,X^2,\ldots,X^n)$ obtained from the naive assumption and making every possible combination of probability distributions on the local credal sets:

$$\mathcal{P}\left(c_k,X^1,\ldots,X^n\right)=\left\{p_c(c_k)\prod_{i=1}^{n}p_{ik}\mid p_c\in\mathcal{P}(C),\,p_{ik}\in\mathcal{P}\left(X^i\mid c_k\right)\right\}. \quad (5)$$

Hence, in order to build the NCC, only the local credal sets are required.

Let $\mathcal{D}$ be a dataset with N instances. Let $n(c_k)$ denote the number of occurences of $C=c_k$ in $\mathcal{D}$, $\forall k=1,2,\ldots,K$. For each $i=1,2\ldots,n$ and for each $1\leq r_i\leq t_i$, let $n\left(x^i_{r_i}\right)$ denote the number of instances in $\mathcal{D}$ for which $X^i=x^i_{r_i}$.

For obtaining the local credal sets, the IDM is usually employed in the literature. According to this model, the credal set on C is the one given by:

$$\mathcal{P}(C)=\left\{p\mid\frac{n(c_k)}{N+s}\leq p(c_k)\leq\frac{n(c_k)+s}{N+s},\quad\forall k=1,2,\ldots,K\right\}. \quad (6)$$

Likewise, under the IDM:

$$\mathcal{P}\left(X^i\mid c_k\right)=\left\{p\mid\frac{n\left(x^i_{r_i,k}\right)}{n(c_k)+s}\leq p(x^i_{r_i}\mid c_k)\leq\frac{n\left(x^i_{r_i,k}\right)+s}{n(c_k)+s},\quad\forall r_i=1,\ldots,t_i\right\}, \quad (7)$$

where $n\left(x^i_{r_i,k}\right)$ is the number of instances in $\mathcal{D}$ that satisfy $\left(C=c_k\wedge X^i=x^i_{r_i}\right)$, $\forall k=1,\ldots,K,\quad\forall i=1,2,\ldots,n$.

For Imprecise Classification, given a new instance for which $X^i = x^i_{r_i}$, with $1 \leq r_i \leq t_i \quad \forall i = 1, 2, \ldots, n$, it is needed to obtain the non-dominated states. For this purpose, it is started from the credal sets defined above.

For each $k = 1, 2, \ldots, K$, $i = 1, 2, \ldots, n$, we denote:

$$\underline{p}(c_k) = \min_{p \in \mathcal{P}(C)} \{p(c_k)\}, \quad \overline{p}(c_k) = \max_{p \in \mathcal{P}(C)} \{p(c_k)\},$$
$$\underline{p}(x^i_{r_i} \mid c_k) = \min_{p_{ik} \in \mathcal{P}(X^i|c_k)} \left\{p_{ik}(x^i_{r_i} \mid c_k)\right\},$$
$$\overline{p}(x^i_{r_i} \mid c_k) = \max_{p_{ik} \in \mathcal{P}(X^i|c_k)} \left\{p_{ik}(x^i_{r_i} \mid c_k)\right\}.$$

It is easy to deduce that, $\forall k = 1, 2, \ldots, K$:

$$\min_{p_c \in \mathcal{P}(C), p_{ik} \in \mathcal{P}(X^i|c_k)} \left\{ p_c(c_k) \prod_{i=1}^{n} p_{ik}(x^i_{r_i} \mid c_k) \right\} = \underline{p}(c_k) \prod_{i=1}^{n} \underline{p}(x^i_{r_i} \mid c_k),$$

$$\max_{p_c \in \mathcal{P}(C), p_{ik} \in \mathcal{P}(X^i|c_k)} \left\{ p_c(c_k) \prod_{i=1}^{n} p_{ik}(x^i_{r_i} \mid c_k) \right\} = \overline{p}(c_k) \prod_{i=1}^{n} \overline{p}(x^i_{r_i} \mid c_k).$$

Since IDM probability intervals are reachable, under this model, the stochastic and credal dominance criteria are equivalent [23]. In consequence, in this case, c_k dominates c_j if, and only if,

$$\underline{p}(c_k) \prod_{i=1}^{n} \underline{p}(x^i_{r_i} \mid c_k) \geq \overline{p}(c_j) \prod_{i=1}^{n} \overline{p}(x^i_{r_i} \mid c_j), \quad \forall j, k \in \{1, 2, \ldots, K\}.$$

Under the IDM:

$$\begin{gathered} \underline{p}(c_k) = \frac{n(c_k)}{N+s}, \quad \overline{p}(c_k) = \frac{n(c_k)+s}{N+s}, \quad \underline{p}(x^i_{r_i} \mid c_k) = \frac{n\left(x^i_{r_i,k}\right)}{n(c_k)+s}, \\ \overline{p}(x^i_{r_i} \mid c_k) = \frac{n\left(x^i_{r_i,k}\right)+s}{n(c_k)+s}, \quad \forall k = 1, 2, \ldots, K. \end{gathered} \tag{8}$$

Consequently, according to NCC, c_k dominates c_j if, and only if:

$$\frac{n(c_k)}{N+s} \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,k}\right)}{n(c_k)+s} \geq \frac{n(c_j)+s}{N+s} \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,j}\right)+s}{n(c_j)+s} \Leftrightarrow$$

$$n\left(c_k\right)\prod_{i=1}^{n}\frac{n\left(x_{r_i,k}^i\right)}{n\left(c_k\right)+s}\geq\left(n\left(c_j\right)+s\right)\prod_{i=1}^{n}\frac{n\left(x_{r_i,j}^i\right)+s}{n\left(c_j\right)+s},\quad\forall j,k\in\left\{1,2,\ldots,K\right\}.$$

Therefore, given a new instance for which $X^i=x_{r_i}^i$, with $1\leq r_i\leq t_i,\quad\forall i=1,2,\ldots,n$, the non-dominated states set predicted by NCC is the one given by:

$$\left\{c_j\mid\left(n\left(c_j\right)+s\right)\prod_{i=1}^{n}\frac{n\left(x_{r_i,j}^i\right)+s}{n\left(c_j\right)+s}>n\left(c_k\right)\prod_{i=1}^{n}\frac{n\left(x_{r_i,k}^i\right)}{n\left(c_k\right)+s},\quad\forall k\neq j\right\},\quad(9)$$

where $j,k\in\{1,2,\ldots,K\}$.

3. Extreme Prior Naive Credal Classifier

Let us assume that the probability for each possible value of the class variable is strictly greater than 0. Our proposed Extreme Prior Naive Credal Classifier (EP-NCC) considers the Bayesian formula with the naive assumption in the following way:

$$\begin{aligned}P\left(C=c_j\mid X^1=x_{r_1}^1,X^2=x_{r_2}^2,\ldots,X^n=x_{r_n}^n\right)\sim\\P\left(C=c_j\right)\prod_{i=1}^{n}P\left(X^i=x_{r_i}^i\mid C=c_j\right)=\\P\left(C=c_j\right)\prod_{i=1}^{n}\frac{P\left(X^i=x_{r_i}^i,C=c_j\right)}{P\left(C=c_j\right)}=\\P\left(C=c_j\right)\prod_{i=1}^{n}\frac{P\left(X^i=x_{r_i}^i\right)P\left(C=c_j\mid X^i=x_{r_i}^i\right)}{P\left(C=c_j\right)}\sim\\P\left(C=c_j\right)\prod_{i=1}^{n}\frac{P\left(C=c_j\mid X^i=x_{r_i}^i\right)}{P\left(C=c_j\right)},\\\forall j=1,\ldots,K,\quad\forall r_i\in\left\{1,2\ldots,t_i\right\},\quad\forall i=1,2,\ldots,n.\end{aligned}\quad(10)$$

We consider the IDM credal set on the class variable C, $\mathcal{P}\left(C\right)$, defined in Equation (6).

Now, for each $i=1,2,\ldots,n$, and for each $r_i\in\{1,2,\ldots,t_i\}$, we consider the following conditional credal set on the class variable:

$$\mathcal{P}\left(C\mid x_{r_i}^i\right)=\left\{p\mid\frac{n\left(x_{r_i,k}^i\right)+s\underline{p}(c_k)}{n\left(x_{r_i}^i\right)+s}\leq p(c_k)\leq\frac{n\left(x_{r_i,k}^i\right)+s\overline{p}(c_k)}{n\left(x_{r_i}^i\right)+s},\quad\forall k=1,\ldots,K\right\},\quad(11)$$

$\underline{p}(c_k)$ and $\overline{p}(c_k)$ being respectively, the minimum and maximum values on c_k among all the probability distributions belonging to $\mathcal{P}(C)$, $\quad \forall k = 1, 2, \ldots, K$. They are obtained via Equation (8).

As NCC, we refer the credal sets $\mathcal{P}(C)$ and $\mathcal{P}\left(C \mid x^i_{r_i}\right)$ as local credal sets, $\forall i = 1, 2, \ldots, n$ $\forall r_i = 1, 2, \ldots, t_i$. The proposed EP-NCC algorithm considers, for each $i = 1, 2, \ldots, n$, and for each $r_i = 1, 2, \ldots, t_i$, the set of joint probability distributions $\mathcal{P}\left(C, x^1_{r_1}, x^2_{r_2}, \ldots, x^n_{r_n}\right)$ resulting from making every possible combination of probability distributions on the local credal sets defined above:

$$\mathcal{P}\left(C, x^1_{r_1}, x^2_{r_2}, \ldots, x^n_{r_n}\right) = \left\{ p_c \prod_{i=1}^{n} \frac{p^i_{r_i}}{p_c}, \mid p_c \in \mathcal{P}(C),\, p_{r_i} \in \mathcal{P}\left(C \mid x^i_{r_i}\right) \right\}, \quad (12)$$

Suppose now that it is wanted to classify a new instance for which $X^i = x^i_{r_i}$, with $1 \leq r_i \leq t_i$.

The following result shows that the local credal sets $\mathcal{P}\left(C \mid x^i_{r_i}\right)$ are defined by reachable probability intervals, $\forall i = 1, 2, \ldots, n$.

Proposition 2. *The set of probability intervals* $I_{C|x^i_{r_i}} = \left\{ \left[\frac{n\left(x^i_{r_i,k}\right) + s\underline{p}(c_k)}{n\left(x^i_{r_i}\right)+s}, \frac{n\left(x^i_{r_i,k}\right) + s\overline{p}(c_k)}{n\left(x^i_{r_i}\right)+s} \right], \quad k = 1, 2, \ldots, K \right\}$ *is reachable,* $\forall i = 1, 2, \ldots, n$.

The proof of this result can be found in Appendix A.

Therefore, in our case, credal and stochastic dominance criteria are equivalent. Consequently, we only analyze the relations among the bounds of the intervals to obtain the non-dominated states set.

We denote

$$\begin{aligned} \underline{p}(c_k \mid x^i_{r_i}) &= \min_{p_{ki} \in \mathcal{P}\left(C|x^i_{r_i}\right)} \left\{ p_{ki}(c_k \mid x^i_{r_i}) \right\}, \\ \overline{p}(c_k \mid x^i_{r_i}) &= \max_{p_{ki} \in \mathcal{P}\left(C|x^i_{r_i}\right)} \left\{ p_{ki}(c_k \mid x^i_{r_i}) \right\}, \end{aligned} \quad (13)$$

$\forall k = 1, 2, \ldots K, \quad \forall i = 1, 2, \ldots, n$, $\mathcal{P}\left(C \mid x^i_{r_i}\right)$ being the credal set defined in Equation (11). Clearly:

$$\underline{p}(c_k \mid x^i_{r_i}) = \frac{n\left(x^i_{r_i,k}\right) + s\underline{p}(c_k)}{n\left(x^i_{r_i}\right) + s}, \quad \overline{p}(c_k \mid x^i_{r_i}) = \frac{n\left(x^i_{r_i,k}\right) + s\overline{p}(c_k)}{n\left(x^i_{r_i}\right) + s},$$

$$\forall k = 1, 2, \ldots K, \quad \forall i = 1, 2, \ldots, n.$$

We may observe that:

$$\min_{p_c\in\mathcal{P}(C),p^i_{r_i}\in\mathcal{P}\left(C|x^i_{r_i}\right)}\left\{p_c(c_k)\prod_{i=1}^{n}\frac{p^i_{r_i}(c_k\mid x^i_{r_i})}{p_c(c_k)}\right\}=\overline{p}(c_k)\prod_{i=1}^{n}\frac{\underline{p}(c_k\mid x^i_{r_i})}{\overline{p}(c_k)}=$$

$$\overline{p}(c_k)\prod_{i=1}^{n}\frac{\frac{n\left(x^i_{r_i,k}\right)+s\underline{p}(c_k)}{n\left(x^i_{r_i}\right)+s}}{\overline{p}(c_k)},\quad\forall k=1,2,\dots K.$$

and

$$\max_{p_c\in\mathcal{P}(C),p^i_{r_i}\in\mathcal{P}\left(C|x^i_{r_i}\right)}\left\{p_c(c_k)\prod_{i=1}^{n}\frac{p^i_{r_i}(c_k\mid x^i_{r_i})}{p_c(c_k)}\right\}=\underline{p}(c_k)\prod_{i=1}^{n}\frac{\overline{p}(c_k\mid x^i_{r_i})}{\underline{p}(c_k)}=$$

$$\underline{p}(c_k)\prod_{i=1}^{n}\frac{\frac{n\left(x^i_{r_i,k}\right)+s\overline{p}(c_k)}{n\left(x^i_{r_i}\right)+s}}{\underline{p}(c_k)},\quad\forall k=1,2,\dots K.$$

Thus, according to EP-NCC, a state c_k dominates another one c_j if:

$$\overline{p}(c_k)\prod_{i=1}^{n}\frac{\frac{n\left(x^i_{r_i,k}\right)+s\underline{p}(c_k)}{n\left(x^i_{r_i}\right)+s}}{\overline{p}(c_k)}\geq\underline{p}(c_j)\prod_{i=1}^{n}\frac{\frac{n\left(x^i_{r_i,j}\right)+s\overline{p}(c_j)}{n\left(x^i_{r_i}\right)+s}}{\underline{p}(c_j)}\Leftrightarrow$$

$$\overline{p}(c_k)\prod_{i=1}^{n}\frac{n\left(x^i_{r_i,k}\right)+s\underline{p}(c_k)}{\overline{p}(c_k)}\geq\underline{p}(c_j)\prod_{i=1}^{n}\frac{n\left(x^i_{r_i,j}\right)+s\overline{p}(c_j)}{\underline{p}(c_j)},$$
$$\forall j,k\in\{1,2,\dots,K\}.$$

Since we are considering the credal set associated with the IDM for C, we have that c_k dominates c_j if, and only if:

$$\frac{n\left(c_k\right)+s}{N+s}\prod_{i=1}^{n}\frac{n\left(x^i_{r_i,k}\right)+s\left(\frac{n(c_k)}{N+s}\right)}{\frac{n(c_k)+s}{N+s}}\geq\frac{n\left(c_j\right)}{N+s}\prod_{i=1}^{n}\frac{n\left(x^i_{r_i,j}\right)+s\left(\frac{n(c_j)+s}{N+s}\right)}{\frac{n(c_j)}{N+s}}\Leftrightarrow$$

$$\left(n\left(c_k\right)+s\right)\prod_{i=1}^{n}\frac{n\left(x^i_{r_i,k}\right)+s\left(\frac{n(c_k)}{N+s}\right)}{n\left(c_k\right)+s}\geq n\left(c_j\right)\prod_{i=1}^{n}\frac{n\left(x^i_{r_i,j}\right)+s\left(\frac{n(c_j)+s}{N+s}\right)}{n\left(c_j\right)},$$
$$\forall j,k\in\{1,2,\dots,K\}.$$

Summarizing, when it is required to classify with EP-NCC a new instance for which $X^i = x^i_{r_i}$, with $1 \leq r_i \leq t_i \quad \forall i = 1, 2, \ldots, n$, the predicted non-dominated states set is given by:

$$\left\{ c_j \mid n\left(c_j\right) \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,j}\right) + s\left(\frac{n(c_j)+s}{N+s}\right)}{n\left(c_j\right)} > \left(n\left(c_k\right)+s\right) \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,k}\right) + s\left(\frac{n(c_k)}{N+s}\right)}{n\left(c_k\right)+s}, \right.$$
$$\left. \forall k \in \{1, 2, \ldots, K\} \setminus \{j\} \right\}. \tag{14}$$

3.1. Justification of the new Naive Credal Classifier

The main difference between the performance of the existing NCC and our proposed EP-NCC resides in the determination of the non-dominated states set.

Suppose that, for an instance that is wanted to be classified, the frequency of a class value is much higher than the class frequency of another class value, the proportion being greater or equal than the proportions of the conditioned class frequencies for all attribute values. Let us also assume that all conditional class frequencies of the latter class value are greater than zero. As we show below, in these situations, if the former class value dominates the latter one under NCC, then the same happens under EP-NCC.

Proposition 3. *Suppose that it is wanted to classify an instance for which* $X^i = x^i_{r_i}$, *with* $1 \leq r_i \leq t_i, \quad \forall i = 1, 2, \ldots, n$. *Let* c_k *and* c_j *be two class values, where* $k, j \in \{1, 2, \ldots, K\}$. *Suppose that* $n\left(x^i_{r_i,j}\right) > 0$ *and* $\frac{n(c_k)}{n(c_j)} \geq \frac{n\left(x^i_{r_i,k}\right)}{n\left(x^i_{r_i,j}\right)}, \quad \forall i = 1, 2, \ldots, n$. *If* c_k *dominates* c_j *under NCC, then* c_k *dominates* c_j *under EP-NCC.*

Proof:. Under our hypothesis of dominance under EP-NCC:

$$n\left(c_k\right) \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,k}\right)}{n\left(c_k\right)+s} \geq \left(n\left(c_j\right)+s\right) \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,j}\right)+s}{n\left(c_j\right)+s} \Rightarrow$$

$$\left(\frac{n\left(c_k\right)}{n\left(c_k\right)+s}\right) \left(\frac{n\left(c_j\right)+s}{n\left(c_k\right)+s}\right)^{n-1} \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,k}\right)}{n\left(x^i_{r_i,j}\right)+s} \geq 1. \tag{15}$$

In addition, by hypothesis, $n\left(x^i_{r_i,j}\right) > 0$ and $\frac{n(c_k)}{n(c_j)} \geq \frac{n\left(x^i_{r_i,k}\right)}{n\left(x^i_{r_i,j}\right)}, \quad \forall i = 1, 2, \ldots, n$. So, it holds that:

$$n(c_k)n\left(x^i_{r_i,j}\right) \geq n(c_j)n\left(x^i_{r_i,k}\right) \Rightarrow \frac{n(c_k)}{N+s}n\left(x^i_{r_i,j}\right) \geq \frac{n(c_j)}{N+s}n\left(x^i_{r_i,k}\right) \Rightarrow$$
$$n\left(x^i_{r_i,j}\right)n\left(x^i_{r_i,k}\right) + \frac{n(c_k)}{N+s}n\left(x^i_{r_i,j}\right) \geq n\left(x^i_{r_i,j}\right)n\left(x^i_{r_i,k}\right) + \frac{n(c_j)}{N+s}n\left(x^i_{r_i,k}\right) \Rightarrow$$
$$\frac{n\left(x^i_{r_i,k}\right) + \frac{n(c_k)}{N+s}}{n\left(x^i_{r_i,j}\right) + \frac{n(c_j)+s}{N+s}} \geq \frac{n\left(x^i_{r_i,k}\right)}{n\left(x^i_{r_i,j}\right)}, \quad \forall i = 1, 2, \ldots, n.$$

250 Thereby:

$$\prod_{i=1}^{n} \frac{n\left(x^i_{r_i,k}\right) + \frac{n(c_k)}{N+s}}{n\left(x^i_{r_i,j}\right) + \frac{n(c_j)+s}{N+s}} \geq \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,k}\right)}{n\left(x^i_{r_i,j}\right)}. \tag{16}$$

Now, since $n\left(x^i_{r_i,j}\right) \leq n\left(c_j\right) \quad \forall i = 1, 2, \ldots, n$:

$$\prod_{i=1}^{n} \frac{n\left(x^i_{r_i,j}\right) + s}{n\left(x^i_{r_i,j}\right)} \geq \prod_{i=1}^{n} \frac{n\left(c_j\right) + s}{n\left(c_j\right)} = \left(\frac{n(c_j)+s}{n(c_j)}\right)^n \geq \left(\frac{n(c_j)+s}{n(c_j)}\right)^{n-1}.$$

In this way:

$$\prod_{i=1}^{n} \frac{n\left(x^i_{r_i,j}\right) + s}{n\left(x^i_{r_i,j}\right)} \geq \left(\frac{n(c_j)+s}{n(c_j)}\right)^{n-1} \Rightarrow$$
$$(n(c_j))^{n-1} \prod_{i=1}^{n} \frac{1}{n\left(x^i_{r_i,j}\right)} \geq (n(c_j)+s)^{n-1} \prod_{i=1}^{n} \frac{1}{n\left(x^i_{r_i,j}\right)+s} \Rightarrow$$
$$\frac{(n(c_j))^{n-1}}{(n(c_k)+s)^{n-1}} \prod_{i=1}^{n} \frac{1}{n\left(x^i_{r_i,j}\right)} \geq \frac{(n(c_j)+s)^{n-1}}{(n(c_k)+s)^{n-1}} \prod_{i=1}^{n} \frac{1}{n\left(x^i_{r_i,j}\right)+s} \Rightarrow$$
$$\frac{(n(c_j))^{n-1}}{(n(c_k)+s)^{n-1}} \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,k}\right)}{n\left(x^i_{r_i,j}\right)} \geq \frac{(n(c_j)+s)^{n-1}}{(n(c_k)+s)^{n-1}} \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,k}\right)}{n\left(x^i_{r_i,j}\right)+s}.$$

251 In consequence:

$$\left(\frac{n\left(c_j\right)}{n\left(c_k\right)+s}\right)^{n-1} \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,k}\right)}{n\left(x^i_{r_i,j}\right)} \geq \left(\frac{n\left(c_j\right)+s}{n\left(c_k\right)+s}\right)^{n-1} \prod_{i=1}^{n} \frac{n\left(x^i_{r_i,k}\right)}{n\left(x^i_{r_i,j}\right)+s}. \tag{17}$$

Hence, we obtain:

$$\left(\frac{n\left(c_j\right)}{n\left(c_k\right)+s}\right)^{n-1}\prod_{i=1}^{n}\frac{n\left(x_{r_i,k}^i\right)+\frac{n(c_j)+s}{N+s}}{n\left(x_{r_i,j}^i\right)+\frac{n(c_k)}{N+s}}\geq$$

Equation (16)

$$\left(\frac{n\left(c_j\right)}{n\left(c_k\right)+s}\right)^{n-1}\prod_{i=1}^{n}\frac{n\left(x_{r_i,k}^i\right)}{n\left(x_{r_i,j}^i\right)}\geq$$

Equation (17)

$$\left(\frac{n\left(c_j\right)+s}{n\left(c_k\right)+s}\right)^{n-1}\prod_{i=1}^{n}\frac{n\left(x_{r_i,k}^i\right)}{n\left(x_{r_i,j}^i\right)+s}\geq$$
$$\left(\frac{n\left(c_k\right)}{n\left(c_k\right)+s}\right)\left(\frac{n\left(c_j\right)+s}{n\left(c_k\right)+s}\right)^{n-1}\prod_{i=1}^{n}\frac{n\left(x_{r_i,k}^i\right)}{n\left(x_{r_i,j}^i\right)+s}\geq 1,$$

where the last inequality is due to Equation (15).

Therefore:

$$\left(\frac{n\left(c_j\right)}{n\left(c_k\right)+s}\right)^{n-1}\prod_{i=1}^{n}\frac{n\left(x_{r_i,k}^i\right)+\frac{n(c_k)}{N+s}}{n\left(x_{r_i,j}^i\right)+\frac{n(c_j)+s}{N+s}}\geq 1\Rightarrow$$
$$\left(\frac{1}{n\left(c_k\right)+s}\right)^{n-1}\prod_{i=1}^{n}\left(n\left(x_{r_i,k}^i\right)+\frac{n\left(c_k\right)}{N+s}\right)\geq$$
$$\left(\frac{1}{n\left(c_j\right)}\right)^{n-1}\prod_{i=1}^{n}\left(n\left(x_{r_i,j}^i\right)+\frac{n\left(c_j\right)+s}{N+s}\right)\Rightarrow$$
$$\left(n\left(c_k\right)+s\right)\prod_{i=1}^{n}\frac{n\left(x_{r_i,k}^i\right)+\frac{n(c_k)}{N+s}}{n\left(c_k\right)+s}\geq n\left(c_j\right)\prod_{i=1}^{n}\frac{n\left(x_{r_i,j}^i\right)+\frac{n(c_j)+s}{N+s}}{n\left(c_j\right)},$$

which implies that c_k dominates c_j under EP-NCC. □

The following example illustrates a case in which, intuitively, it makes much sense that a state dominates another one, and such a dominance case is verified with EP-NCC but not with NCC.

Example 1. *Let $\mathcal{D}$ be a dataset with 50 instances and two attributes X^1 and X^2. Let us assume that each attribute X^i takes values on $\{x_1^i, x_2^i\}$, for $i = 1, 2$. Let C be the class variable whose set of possible values is $\{c_1, c_2\}$. Suppose that $n(c_1) = 48$ and $n(c_2) = 2$. Let us assume the following arrangement of the instances for each one of the possible values of the attributes:*

$$
\begin{aligned}
&X^1 = x_1^1 \rightarrow n\left(x_{1,1}^1\right) = 48, \quad n\left(x_{1,2}^1\right) = 1.\\
&X^1 = x_2^1 \rightarrow n\left(x_{2,1}^1\right) = 0, \quad n\left(x_{2,2}^1\right) = 1.\\
&X^2 = x_1^2 \rightarrow n\left(x_{1,1}^2\right) = 15, \quad n\left(x_{1,2}^2\right) = 1.\\
&X^2 = x_2^2 \rightarrow n\left(x_{2,1}^2\right) = 33, \quad n\left(x_{2,2}^2\right) = 1.
\end{aligned}
$$

We assume the value $s = 1$ for the IDM hyperparameter. Suppose that it is required to classify a new instance for which $X^1 = x_2^1$ and $X^2 = x_2^2$. Then:

$$
\begin{aligned}
&(n(c_1) + 1) \times \frac{n\left(x_{2,1}^1\right) + \left(\frac{n(c_1)}{N+1}\right)}{n(c_1) + 1} \times \frac{n\left(x_{2,1}^2\right) + \left(\frac{n(c_1)}{N+1}\right)}{n(c_1) + 1} =\\
&49 \times \frac{\frac{48}{51}}{49} \times \frac{33 + \frac{48}{51}}{49} = 0.6519 > 0.5605 = 2 \times \frac{1 + \frac{3}{51}}{2} \times \frac{1 + \frac{3}{51}}{2} =\\
&n(c_2) \times \frac{n\left(x_{2,2}^1\right) + \left(\frac{n(c_2)+1}{N+1}\right)}{n(c_2)} \times \frac{n\left(x_{2,2}^2\right) + \left(\frac{n(c_2)+1}{N+1}\right)}{n(c_2)}.
\end{aligned}
$$

Hence, c_1 dominates c_2 under EP-NCC. However, it does not happen under NCC since:

$$
\begin{aligned}
&n(c_1) \times \frac{n\left(x_{2,1}^1\right)}{n(c_1) + 1} \times \frac{n\left(x_{2,1}^2\right)}{n(c_1) + 1} = 48 \times 0 \times \frac{33}{49} = 0 <\\
&(n(c_2) + 1) \times \frac{n\left(x_{2,2}^1\right) + 1}{n(c_2) + 1} \times \frac{n\left(x_{2,2}^2\right) + 1}{n(c_2) + 1} = 3 \times \frac{2}{3} \times \frac{2}{3} = \frac{4}{3}.
\end{aligned}
$$

In this case, the prediction made by EP-NCC about the dominance of c_1 on c_2 is intuitively more coherent than the one made by NCC.

Therefore, the predictions of EP-NCC are more informative than the predictions of NCC. It is such an important issue for our contribution. In the previous example, we have observed a problematic situation of NCC: when only one lower conditional probability of a class value is equal to 0, the lower probability predicted by NCC for that state of the class variable

is equal to 0, even though the rest of the lower conditional probabilities are very high. Thus, this state does not dominate any other only due to that conditional lower probability, which is incoherent. This problem is solved with our proposed EP-NCC because it considers the lower prior probabilities of the class values in the estimation of the lower conditional probabilities.

Our proposal assumes more risk of making incorrect predictions than NCC since its predicted non-dominated states set is often smaller. However, this risk is controlled because the probability interval predicted for each class value is narrowed by taking the extreme prior probabilities into account. In addition, EP-NCC solves some problematics situations of NCC, such as the one found in Example 1.

To summarize, EP-NCC is more appropriate than NCC for Imprecise Classification since the former method is more informative than the latter without assuming a much higher risk of making erroneous predictions. This fact is corroborated in Section 4 with an exhaustive experimental analysis.

4. Experimentation

In this experimental research, we aim to compare the performance of the existing NCC, our proposed EP-NCC, and the Imprecise Credal Decision Tree algorithm (ICDT).

4.1. Experimental setup

4.1.1. Datasets

The three algorithms considered in this experimentation have been applied to 34 known classification datasets. These datasets are different in terms of sample size, number of continuous and discrete attributes, number of values per feature, number of values of the class variable, etc. The datasets have been chosen in such a way that they have at least 3 class values since, with only two states of the class variable, an algorithm for Imprecise Classification always predicts all of them or only one. These datasets can be found in the *UCI Machine Learning repository* [24]. Table 1 shows the most important characteristics of each dataset.

4.1.2. Evaluation metrics

Consistently with [5], we evaluate the performance of the classifiers using the two following evaluation metrics for Imprecise Classification:

Table 1: Dataset description. Column "N" is the number of instances in the dataset, column "Attr" is the number of attributes, column "Cont" is the number of continuous variables, column "Disc" is the number of discrete features, column "K" is the number of class values, and column "Range" is the range of values of the discrete variables of each dataset.

Dataset	N	Attr	Cont	Disc	K	Range
anneal	898	38	6	32	6	2-10
arrhythmia	452	279	206	73	16	2
audiology	226	69	0	69	24	2-6
autos	205	25	15	10	7	2-22
balance-scale	625	4	4	0	3	-
car	1728	6	0	6	4	3-4
cmc	1473	9	2	7	3	2-4
dermatology	366	34	1	33	6	2-4
ecoli	366	7	7	0	7	-
flags	194	30	2	28	8	2-13
hypothyroid	3772	30	7	23	4	2-4
iris	150	4	4	0	3	-
letter	20000	16	16	0	26	-
lymphography	146	18	3	15	4	2-8
mfeat-pixel	2000	240	0	240	10	4-6
nursery	12960	8	0	8	4	2-4
optdigits	5620	64	64	0	10	-
page-blocks	5473	10	10	0	5	-
pendigits	10992	16	16	0	10	-
postop-patient-data	90	9	0	9	3	2-4
primary-tumor	339	17	0	17	21	2-3
segment	2310	19	16	0	7	-
soybean	683	35	0	35	19	2-7
spectrometer	531	101	100	1	48	4
splice	3190	60	0	60	3	4-6
sponge	76	44	0	44	3	2-9
tae	151	5	3	2	3	2
vehicle	946	18	18	0	4	-
vowel	990	11	10	1	11	2
waveform	5000	40	40	0	3	-
wine	178	13	13	0	3	-
zoo	101	16	1	16	7	2

- **Discounted Accuracy measure** (DACC) [9]. Let N_{Test} denote the number of instances of the test set, U_i the predicted states set for the ith instance, and $|U_i|$ its cardinality. DACC is defined as follows:

$$DACC = \frac{1}{N_{Test}} \sum_{i=1}^{N_{Test}} \frac{(correct)_i}{|U_i|}, \tag{18}$$

  where $(correct)_i$ is equal to 1 if U_i contains the real value of the class variable and 0 else, $\forall i = 1, 2, \ldots, N_{Test}$.

- **MIC** [5]: It is defined in the following way:

$$MIC = \frac{1}{N_{Test}} \left(\sum_{i:Success} \log \frac{|U_i|}{k} + \frac{1}{k-1} \sum_{i:Error} \log k \right). \tag{19}$$

Furthermore, we use the following measures to separately evaluate how informative are the imprecise classifiers considered in this experimentation, and their performance in making right predictions:

- **Determinacy**: It indicates the proportion of instances for which a single state of the class variable is predicted.

- **Single Accuracy**: The accuracy among the instances for which a single class value is predicted.

- **Set Accuracy**: It measures, among the instances for which there are at least two non-dominated states, the proportion of instances for which the real class value is among the predicted ones.

- **Indeterminacy size**: The average size of the non-dominated states set.

4.1.3. Procedure

We have used the Weka software [25] for this experimentation. We have added the necessary structures and methods for using the three algorithms considered here. Regarding the IDM hyperparameter, remark that two values are recommended by Walley in [3]: $s = 1$ and $s = 2$. We have used the value $s = 1$ because it produces more precise predictions than $s = 2$, requires less

computational cost and is the one employed in other experimental analyses carried out in the literature about Imprecise Classification algorithms.

The datasets have been preprocessed as in [5]: missing values have been replaced with mean values for continuous attributes and modal values for discrete variables. After that, continuous attributes have been discretized following the Fayyad and Irani's discretization method [26]. We have employed the Weka's filters for the preprocessing. The preprocessing has been applied to the training set and then, it has been translated to the test set.

For each dataset and each algorithm considered in this experimental study, a 10 times 10-fold cross-validation procedure has been repeated.

4.1.4. Statistical Evaluation

Following the recommendations given in [27] for statistical comparisons among three or more algorithms, we have compared the results obtained by NCC, EP-NCC, and ICDT via the following statistical tests with a level of significance of $\alpha = 0.05$:

- **Friedman test** [28]: It is a non-parametric test that ranks the obtained results separately for each dataset (the algorithm that attains the best result is assigned to rank 1, the one that obtains the second-best result to rank 2, etc). The null hypothesis of this test is that all the algorithms perform equivalently.

- If the null hypothesis of the Friedman test is rejected, the algorithms are compared to each other via the **Nemenyi test** [29].

4.2. Results and discussion

Tables 2 and 3 allow us to see a summary of the results obtained in the DACC and MIC measures, respectively. Specifically, they show the average values and the average Friedman ranks obtained by each classifier in each one of these metrics, as well as the cases of statistically significant differences according to the Nemenyi test. The best results in average values and Friedman ranks are marked in bold. The complete results of these metrics can be found in Appendix B.

It can be observed that both ICDT and EP-NCC perform significantly better than NCC in DACC and MIC. In fact, there are three algorithms, and the Friedman rank obtained by NCC is higher than 2.6 in both DACC and MIC, which implies that NCC obtains the worst results in most of the

datasets. In both evaluation metrics, EP-NCC and ICDT perform equivalently according to the Nemenyi test, although the average values are higher for EP-NCC, and the Friedman ranks achieved by EP-NCC are lower than the ones obtained by ICDT.

Algorithm	Average	Friedman rank	Wins Nemenyi
NCC	0.6237	2.6765	
EP-NCC	**0.7810**	**1.5882**	NCC
ICDT	0.7763	1.7353	NCC

Table 2: Summary of DACC results. The column "Wins Nemenyi" indicates the algorithms outperformed by the one in the row via the Nemenyi test.

Algorithm	Average	Friedman rank	Wins Nemenyi
NCC	1.3042	2.8235	
EP-NCC	**1.3529**	**1.5882**	NCC
ICDT	1.7676	1.8235	NCC

Table 3: Summary of MIC results. The column "Wins Nemenyi" indicates the algorithms outperformed by the one in the row via the Nemenyi test.

For a deeper analysis, Tables 4 and 5 illustrate, respectively, the average values and average Friedman ranks obtained by each classifier in Determinacy, Single Accuracy, Set Accuracy, and Indeterminacy size. The cases of statistically significant differences in each one of these metrics according to the Nemenyi test can be seen in Table 6. Again, the best results are marked in bold. In Appendix B, the complete results of these metrics can be found.

Metric	NCC	EP-NCC	ICDT
Determinacy	0.6037	0.9150	**0.9477**
Single Accuracy	**0.8692**	0.8165	0.8058
Indeterminacy size	4.4537	**2.0999**	5.3294
Set Accuracy	0.8698	0.8389	**0.8999**

Table 4: Average results in Determinacy, Single Accuracy, Indeterminacy size and Set Accuracy

We remark the following points about the results obtained in these particular measures:

Metric	NCC	EP-NCC	ICDT
Determinacy	2.8235	**1.5**	1.6765
Single Accuracy	**1.3226**	2.3548	2.3226
Indeterminacy size	2.1875	**1.0469**	2.7656
Set Accuracy	1.9844	2.3594	**1.6562**

Table 5: Average Friedman ranks in Determinacy, Single Accuracy, Indeterminacy size and Set Accuracy.

	NCC	EP-NCC	ICDT
NCC	-	Determinacy, Indeterminacy size	Determinacy
EP-NCC	Single Accuracy	-	Set Accuracy
ICDT	Single Accuracy	Indeterminacy size	-

Table 6: Results of the Nemenyi test in Determinacy, Single Accuracy, Indeterminacy size and Set Accuracy. The algorithm in each column performs significantly better than the one in the row in the metrics indicated in the cell.

- NCC obtains, by far, the worst performance in predicting a single state of the class variable due to the results obtained in **Determinacy**. In this metric, EP-NCC and ICDT perform equivalently; EP-NCC obtains a lower Friedman rank than ICDT, whereas the second algorithm obtains a higher average value than the first one; but there are no statistically significant differences according to the Nemenyi test.

- Regarding **Single Accuracy**, which measures the accuracy among the instances for which a single class value is predicted, NCC obtains significantly better results than the other two algorithms. Nevertheless, we should remark that, with NCC, the percentage of instances precisely classified is far lower than with the other two algorithms. Again, there are no statistically significant differences between EP-NCC and ICDT.

- The results obtained in **Indeterminacy size** allow deducing that EP-NCC achieves, by far, the lowest average number of non-dominated states. It performs significantly better than NCC and ICDT according to the Friedman and Nemenyi tests in this metric. There are no statistically significant differences between NCC and ICDT in this measure, even though NCC attains a lower average value than ICDT, and the Friedman rank of NCC is also lower.

- Concerning **Set Accuracy**, which measures the percentage of correct

predictions among the instances for which more than a class value is predicted, ICDT obtains the best results according to Friedman rank and average values. Furthermore, ICDT significantly outperforms EP-NCC via the Nemenyi test. There are no statistically significant differences between NCC and EP-NCC, although the first algorithm obtains a higher average value and a lower Friedman rank than the second one.

Also, we aim to compare the computational complexity of the algorithms considered in this experimentation. For this purpose, Table 7 shows a summary of the processing time, in milliseconds, of NCC, EP-NCC, and ICDT. The complete processing time results can be found in Appendix B.

Algorithm	Average	Friedman rank	Wins Nemenyi
NCC	0.0006	1.5588	ICDT
EP-NCC	**0.0005**	**1.4412**	ICDT
ICDT	0.0424	3	

Table 7: Summary of processing time results. The column "Wins Nemenyi" indicates the algorithms outperformed by the one in the row via the Nemenyi test.

We can see that ICDT requires, by far, the highest computational complexity. Indeed, it obtains the worst result in terms of processing time in all the datasets since the average Friedman rank is equal to 3. In addition, the average processing time achieved by ICDT is much higher than the ones obtained by NCC and EP-NCC. NCC obtains a higher average value than EP-NCC, and the Friedman rank achieved by NCC is also higher. However, there are no statistically significant differences between these two algorithms in terms of processing time according to the Nemenyi test.

Summary of the results:. It can be observed that EP-NCC outperforms NCC. More specifically, EP-NCC performs better than NCC in predicting fewer class values and, thus, EP-NCC is more informative than NCC, as we argued in Section 3.1. According to the results obtained in Single Accuracy and Set Accuracy, NCC performs slightly better than EP-NCC in making the right predictions. It is because the predictions made by EP-NCC are more precise than the ones made by NCC and, consequently, the first method assumes more risk of making incorrect predictions than the second one. Nonetheless, this is not as important as the fact that EP-NCC is more informative since the risk is not far higher, as we commented in Section 3, and it is shown in the results obtained in MIC and DACC.

Moreover, ICDT and EP-NCC perform equivalently. Both algorithms obtain statistically equivalent results in predicting only one state of class variable and in the accuracy among precise predictions. Regarding instances imprecisely classified, ICDT outperforms EP-NCC in making the right predictions, while EP-NCC performs significantly better than ICDT in predicting fewer states of the class variable.

The processing times of NCC and EP-NCC are very similar. The ICDT algorithm requires a much higher computational complexity than NCC and EP-NCC.

5. Conclusions

In this research, we have proposed a new version of the Naive Credal Classifier, called the Extreme Prior Naive Credal Classifier. It is based on the already existing Naive Bayes algorithm for Imprecise Classification, but it takes into consideration the lower and upper prior probabilities of the class variable in the estimation of the lower and upper conditional probabilities, unlike the original Naive Credal Classifier.

We have shown that our proposed EP-NCC algorithm tends to predict fewer states of the class variable than NCC. In consequence, the predictions made by our proposed method are more informative. It also implies that, with EP-NCC, there is more risk of making erroneous predictions. Nevertheless, as we have argued, this risk is not far higher than with NCC since EP-NCC narrows the conditioned probability intervals by considering the bounds of the prior probabilities of the class variable.

An exhaustive experimental study has been carried out in this work to compare the performance of the existing NCC, our proposed EP-NCC, and the Imprecise Credal Decision Tree method. This experimental analysis has shown that EP-NCC significantly outperforms NCC. More specifically, the predictions made by EP-NCC are far more informative than the ones made by NCC, whereas the difference between both algorithms in making correct predictions is not statistically significant. In addition, ICDT and EP-NCC perform equivalently; while ICDT obtains better results than EP-NCC in metrics corresponding to Accuracy, the predictions made by EP-NCC are more informative. Nonetheless, ICDT requires a considerably higher computational complexity than EP-NCC.

For the reasons above exposed, good performance, and low computational cost, it can be concluded that our proposed EP-NCC algorithm is more

suitable to be applied on large datasets for Imprecise Classification than the existing algorithms for such a task. Due to the increasing amount of data used in every area, these characteristics are very important points to take into account in favor of our proposal.

Appendix A. Proof of Proposition 2

Proof:. (Proposition 2)

For each $i = 1, 2, \ldots, n$ we have that:

$$\frac{n\left(x^i_{r_i,j}\right) + s\underline{p}(c_j)}{n\left(x^i_{r_i}\right) + s} + \sum_{k=1,k\neq j}^{K} \frac{n\left(x^i_{r_i,k}\right) + s\overline{p}(c_k)}{n\left(x^i_{r_i}\right) + s} =$$
$$\sum_{k=1}^{K} \frac{n\left(x^i_{r_i,k}\right)}{n\left(x^i_{r_i}\right) + s} + \frac{s}{n\left(x^i_{r_i}\right) + s}\left(\underline{p}(c_j) + \sum_{k=1,k\neq j}^{K} \overline{p}(c_k)\right) \geq$$
$$\frac{n\left(x^i_{r_i}\right)}{n\left(x^i_{r_i}\right) + s} + \frac{s}{n\left(x^i_{r_i}\right) + s} = 1, \quad \forall j = 1, 2, \ldots, K.$$

since sets of IDM probability intervals are reachable.

Analogously, it can be checked that

$$\frac{n\left(x^i_{r_i,j}\right) + s\overline{p}(c_j)}{n\left(x^i_{r_i}\right) + s} + \sum_{k=1,k\neq j}^{K} \frac{n\left(x^i_{r_i,k}\right) + s\underline{p}(c_k)}{n\left(x^i_{r_i}\right) + s} \leq 1, \quad \forall j = 1, 2, \ldots, K,$$

and Proposition 1 allows us to conclude that $I_{C|x^i_{r_i}}$ is reachable.

□

Appendix B. Complete results from the experimentation

In this Section, the complete results of the experimentation are shown. We mark the best results in bold.

Table 8: Complete DACC results

Dataset	NCC	EP-NCC	ICDT
anneal	0.4636	0.9793	**0.9957**
arrhythmia	0.1480	**0.6756**	0.6625
audiology	0.1190	0.6926	**0.7887**
autos	0.2780	0.6935	**0.7817**
balance-scale	**0.7220**	0.7170	0.6961
bridges-version1	0.3112	0.6339	**0.6375**
bridges-version2	0.3097	**0.6456**	0.5729
car	0.8575	0.8566	**0.9168**
cmc	**0.5015**	0.5001	0.4884
dermatology	0.8764	**0.9845**	0.9405
ecoli	0.8022	**0.8154**	0.7993
flags	0.1250	**0.5934**	0.5554
hypothyroid	0.9257	0.9882	**0.9935**
iris	0.9287	0.9330	**0.9337**
letter	0.7431	0.7478	**0.7714**
lymphography	0.6615	**0.8453**	0.7275
mfeat-pixel	0.7157	**0.9376**	0.7702
nursery	0.9027	0.9031	**0.9628**
optdigits	0.9192	**0.9223**	0.7716
page-blocks	0.9336	0.9392	**0.9619**
pendigits	0.8798	**0.8814**	0.8812
postoperative-patient-data	0.4828	0.6515	**0.7104**
primary-tumor	0.1539	**0.4500**	0.3815
segment	0.9092	0.9189	**0.9406**
soybean	0.8973	**0.9386**	0.9178
spectrometer	0.1467	**0.4489**	0.4430
splic	0.9526	**0.9553**	0.9270
sponge	0.3593	**0.9410**	0.9293
tae	0.4468	0.4459	**0.4678**
vehicle	0.6056	0.6104	**0.6899**
vowel	0.5640	0.5902	**0.7635**
waveform	0.7996	**0.7997**	0.7371
wine	0.9378	**0.9840**	0.9194
zoo	0.8272	0.9347	**0.9592**
Average	0.6237	**0.7810**	0.7763

Table 9: Complete MIC results

Dataset	NCC	EP-NCC	ICDT
anneal	1.0219	1.7811	**1.7889**
arrhythmia	0.9809	2.5301	**2.6901**
audiology	0.6736	2.8356	**3.1017**
autos	0.8748	**1.8164**	1.8021
balance-scale	0.9365	**0.9767**	0.9631
bridges-version1	0.4827	**1.5929**	1.5794
bridges-version2	0.4751	**1.6081**	1.4061
car	1.3168	1.3314	**1.3464**
cmc	0.8992	**0.9106**	0.8763
dermatology	1.5451	**1.7797**	1.7478
ecoli	1.9154	**2.0086**	1.9977
flags	0.0000	**1.9223**	1.7227
hypothyroid	1.2899	1.3816	**1.3835**
iris	1.0551	**1.0731**	1.0672
letter	3.2028	**3.2249**	3.0505
lymphography	0.9886	**1.3083**	1.2623
mfeat-pixel	1.7206	**2.2854**	2.0326
nursery	1.5765	1.5779	**1.5959**
optdigits	2.2685	**2.2832**	2.0883
page-blocks	1.5708	1.5893	**1.5914**
pendigits	2.2627	**2.2750**	2.1863
postoperative-patient-data	0.5699	0.9134	**0.9916**
primary-tumor	1.8127	2.8668	**2.9242**
segment	1.8612	**1.9226**	1.8996
soybean	2.8220	**2.9264**	2.9161
spectrometer	1.5155	**3.6503**	3.5774
splice	1.0721	**1.0810**	1.0538
sponge	0.0631	**1.0678**	1.0561
tae	0.6435	0.7648	**0.9031**
vehicle	1.2039	1.2390	**1.2614**
vowel	2.0967	**2.2723**	2.2648
waveform	1.0208	**1.0243**	0.9751
wine	1.0021	**1.0913**	1.0601
zoo	1.6012	1.8762	**1.9346**
Average	1.3042	**1.7879**	1.7676

Table 10: Complete Determinacy results

Dataset	NCC	EP-NCC	ICDT
anneal	0.0031	0.9916	**0.9990**
arrhythmia	0.0151	0.7026	**0.9429**
audiology	0.0027	0.6105	**0.9454**
autos	0.0000	0.8976	**0.9454**
balance-scale	0.8543	**0.9567**	0.9401
bridges-version1	0.1516	0.7890	**0.9070**
bridges-version2	0.1442	**0.8389**	0.8263
car	0.9609	**0.9893**	0.9889
cmc	0.9608	**0.9878**	0.9322
dermatology	0.8405	**0.9863**	0.9830
ecoli	0.8626	0.9539	**0.9620**
flags	0.0000	**0.8857**	0.8524
hypothyroid	0.8682	0.9988	**0.9995**
iris	0.9673	**0.9980**	0.9860
letter	0.9469	**0.9957**	0.8997
lymphography	0.5305	0.9523	**0.9665**
mfeat-pixel	0.6452	**0.9938**	0.8830
nursery	0.9965	**0.9994**	0.9980
optdigits	0.9732	**0.9969**	0.9008
page-blocks	0.9717	**0.9987**	0.9950
pendigits	0.9782	**0.9994**	0.9487
postoperative-patient-data	0.4111	0.8878	**0.9989**
primary-tumor	0.0257	0.6519	**0.8009**
segment	0.9365	**0.9984**	0.9811
soybean	0.9060	0.9829	**0.9834**
spectrometer	0.0379	0.6348	**0.8271**
splice	0.9830	**0.9973**	0.9788
sponge	0.0000	0.9695	**0.9805**
tae	0.3546	0.6697	**0.9973**
vehicle	0.8940	0.9714	**0.9723**
vowel	0.6535	0.9186	**0.9417**
waveform	0.9914	**0.9980**	0.9672
wine	0.8974	**0.9972**	0.9893
zoo	0.7597	0.9109	**1.0000**
Average	0.6037	0.9150	**0.9477**

Table 11: Complete Indeterminacy size results

Dataset	NCC	EP-NCC	ICDT
anneal	2.2251	**2.0000**	6.0000
arrhythmia	8.4307	**2.5549**	3.5072
audiology	15.4544	**3.0851**	4.8801
autos	3.5564	**2.0425**	6.5304
balance-scale	**2.0000**	**2.0000**	2.1149
bridges-version1	5.0785	**2.1801**	4.5417
bridges-version2	5.0677	**2.2303**	5.8016
car	2.0381	**2.0115**	3.6667
cmc	2.0036	**2.0000**	2.3263
dermatology	5.0643	**2.0000**	6.0000
ecoli	3.2957	**2.0068**	3.8911
flags	8.0000	**2.1073**	7.3432
hypothyroid	2.0927	**2.0000**	4.0000
iris	2.0303	**2.0000**	**2.0000**
letter	2.4407	**2.0075**	14.1980
lymphography	2.5067	**2.0196**	3.2276
mfeat-pixel	6.2712	**2.0143**	8.9172
nursery	2.0158	**2.0000**	2.8688
optdigits	2.3625	**2.0000**	7.8818
page-blocks	2.6810	**2.0000**	4.5360
pendigits	2.4041	**2.0000**	8.2314
postoperative-patient-data	2.3641	**2.0156**	3.0000
primary-tumor	7.0683	**2.5082**	3.3272
segment	3.6898	**2.0000**	6.5570
soybean	5.6971	**2.0000**	6.4303
spectrometer	24.4356	**2.5013**	17.1857
splice	2.1977	**2.0000**	2.9897
sponge	2.8443	**2.0000**	3.0000
tae	2.0000	2.0000	**1**
vehicle	2.1226	**2.0112**	2.8442
vowel	2.5287	**2.0487**	6.8725
waveform	**2.0000**	**2.0000**	2.9429
wine	2.6023	**2.0000**	2.9286
zoo	4.8565	2.0515	**1**
Average	4.4537	**2.0999**	5.3294

Table 12: Complete Single Accuracy results

Dataset	NCC	EP-NCC	ICDT
anneal	**1.0000**	0.9821	0.9970
arrhythmia	**0.9524**	0.8448	0.7028
audiology	**1.0000**	0.9105	0.8656
balance-scale	**0.7815**	0.7317	0.7187
bridges-version1	**0.9390**	0.7191	0.6797
bridges-version2	**0.9618**	0.7034	0.6604
car	0.8730	0.8608	**0.9237**
cmc	**0.5049**	0.5014	0.4983
dermatology	**0.9968**	0.9912	0.9539
ecoli	**0.8734**	0.8341	0.8196
hypothyroid	**0.9943**	0.9888	0.9939
iris	**0.9434**	0.9338	0.9401
letter	0.7700	0.7498	**0.8402**
lymphography	**0.9305**	0.8638	0.7412
mfeat-pixel	**0.9875**	0.9411	0.8564
nursery	0.9044	0.9033	**0.9639**
optdigits	**0.9341**	0.9239	0.8398
page-blocks	0.9515	0.9398	**0.9654**
pendigits	0.8914	0.8816	**0.9197**
postoperative-patient-data	0.6143	0.6702	**0.7093**
primary-tumor	**0.7769**	0.5702	0.4581
segment	0.9476	0.9197	**0.9553**
soybean	**0.9567**	0.9479	0.9265
spectrometer	**0.7616**	0.5298	0.5128
splice	**0.9615**	0.9567	0.9399
tae	**0.6236**	0.5200	0.4697
vehicle	0.6355	0.6165	**0.6988**
vowel	0.6770	0.6103	**0.7953**
waveform	**0.8022**	0.8003	0.7504
wine	**0.9994**	0.9853	0.9255
zoo	**1.0000**	0.9788	0.9592
Average	**0.8692**	0.8165	0.8058

Table 13: Complete Set Accuracy results

Dataset	NCC	EP-NCC	ICDT
anneal	0.9892	0.9815	**1.0000**
arrhythmia	**0.8937**	0.7811	0.3292
audiology	**1.0000**	0.8734	0.8284
autos	0.8558	0.7503	**0.9868**
balance-scale	0.8176	**0.8432**	0.7518
bridges-version1	0.9472	0.7048	**0.9673**
bridges-version2	0.9446	0.7222	**0.9851**
car	0.9646	0.9109	**0.9973**
cmc	**0.8363**	0.8060	0.8052
dermatology	**1.0000**	**1.0000**	**1.0000**
ecoli	**0.9280**	0.8534	0.8535
flags	**1.0000**	0.7359	0.9809
hypothyroid	0.9742	**1.0000**	**1.0000**
iris	**1.0000**	**1.0000**	**1.0000**
letter	0.5731	0.6052	**0.7926**
lymphography	0.8728	0.9739	**1.0000**
mfeat-pixel	**0.9667**	0.7495	0.9340
nursery	0.8966	**1.0000**	**1.0000**
optdigits	0.8356	0.8094	**0.8922**
page-blocks	0.7852	0.8794	**0.9931**
pendigits	0.7923	0.8654	**0.9603**
postoperative-patient-data	0.8864	**1.0000**	**1.0000**
primary-tumor	**0.6967**	0.5256	0.3530
segment	0.9558	0.8929	**0.9920**
soybean	0.9277	0.8072	**0.9788**
spectrometer	**0.8988**	0.6117	0.6035
splice	0.9405	0.9226	**1.0000**
sponge	**1.0000**	0.5614	**1.0000**
tae	0.7719	**0.7815**	
vehicle	0.7440	0.8372	**0.8734**
vowel	0.7953	0.7355	**0.9387**
waveform	**1.0000**	**1.0000**	**1.0000**
wine	**1.0000**	**1.0000**	**1.0000**
zoo	**1.0000**	**1.0000**	
Average	0.8968	0.8389	**0.8999**

Table 14: Complete processing time results

Dataset	NCC	EP-NCC	ICDT
anneal	0.0014	**0.0009**	0.0112
arrhythmia	**0.0009**	0.0014	0.0916
audiology	**0.0001**	**0.0001**	0.0134
autos	**0.0001**	**0.0001**	0.0040
balance-scale	**0.0001**	**0.0001**	0.0009
bridges-version1	**0.0001**	0.0002	0.0007
bridges-version2	**0.0001**	**0.0001**	0.0011
car	0.0003	**0.0001**	0.0020
cmc	**0.0001**	**0.0001**	0.0043
dermatology	0.0003	**0.0001**	0.0039
ecoli	0.0002	**0.0001**	0.0008
flags	0.0002	**0.0001**	0.0056
hypothyroid	**0.0006**	0.0007	0.0090
iris	**0.0001**	**0.0001**	0.0002
letter	0.0034	**0.0026**	0.3676
lymphography	**0.0001**	0.0002	0.0006
mfeat-pixel	**0.0037**	**0.0037**	0.3366
nursery	0.0014	**0.0012**	0.0328
optdigits	**0.0020**	0.0021	0.2257
page-blocks	**0.0004**	0.0006	0.0102
pendigits	0.0016	**0.0015**	0.0938
postoperative-patient-data	**0.0001**	**0.0001**	0.0002
primary-tumor	0.0001	**0.0000**	0.0065
segment	**0.0002**	**0.0002**	0.0114
soybean	0.0002	**0.0001**	0.0115
spectrometer	0.0005	**0.0003**	0.0919
splice	**0.0010**	0.0012	0.0478
sponge	0.0002	**0.0001**	0.0006
tae	**0.0001**	**0.0001**	0.0002
vehicle	0.0003	**0.0001**	0.0041
vowel	**0.0001**	0.0003	0.0071
waveform	0.0010	**0.0008**	0.0436
wine	**0.0001**	0.0002	0.0003
zoo	**0.0001**	**0.0001**	0.0004
Average	0.0006	**0.0005**	0.0424

References

[1] D. J. Hand, Construction and Assessment of Classification Rules, John Wiley and Sons, New York, 1997.

[2] M. Zaffalon, The naive credal classifier, Journal of Statistical Planning and Inference 105 (1) (2002) 5 – 21, imprecise Probability Models and their Applications. doi:10.1016/S0378-3758(01)00201-4.

[3] P. Walley, Inferences from multinomial data; learning about a bag of marbles (with discussion), Journal of the Royal Statistical Society. Series B (Methodological) 58 (1) (1996) 3–57. doi:10.2307/2346164.

[4] L. M. De Campos, J. F. Huete, S. Moral, Probability intervals: a tool for uncertain reasoning, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 02 (02) (1994) 167–196. doi:10.1142/S0218488594000146.

[5] J. Abellán, A. R. Masegosa, Imprecise classification with credal decision trees, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 20 (05) (2012) 763–787. doi:10.1142/S0218488512500353.

[6] G. Corani, C. de Campos, A tree augmented classifier based on extreme imprecise dirichlet model, International Journal of Approximate Reasoning 51 (9) (2010) 1053 – 1068, imprecise probability in statistical inference and decision making. doi:10.1016/j.ijar.2010.08.007.

[7] A. Antonucci, G. Corani, The multilabel naive credal classifier, International Journal of Approximate Reasoning 83 (2017) 320 – 336. doi:10.1016/j.ijar.2016.10.006.

[8] M. Zaffalon, Credible classification for environmental problems, Environmental Modelling & Software 20 (8) (2005) 1003 – 1012, methods of Uncertainty Treatment in Environmental Models. doi:10.1016/j.envsoft.2004.10.006.

[9] G. Corani, M. Zaffalon, Learning reliable classifiers from small or incomplete data sets: the naive credal classifier 2, Journal of Machine Learing Research 9 (2008) 581–621.

[10] R. O. Duda, P. E. Hart, Pattern Classification and Scene Analysis, John Wiley and Sons, New York, 1973.

[11] P. Domingos, M. Pazzani, On the optimality of the simple bayesian classifier under zero-one loss, Machine Learning 29 (2) (1997) 103–130. doi:10.1023/A:1007413511361.

[12] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, Machine Learning 29 (2) (1997) 131–163. doi:10.1023/A:1007465528199.

[13] P. Langley, W. Iba, K. Thompson, An analysis of bayesian classifiers, in: Proceedings of the 10th national conference on Artificial intelligence (AAAI'92), MIT Press, 1992, pp. 223–228.

[14] J. L. Hellerstein, T. S. Jayram, I. Rish, Recognizing end-user transactions in performance management, in: Proceedings of the 7th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence, 2000, pp. 596–602.

[15] H. Zhang, D. Li, Naïve bayes text classifier, in: IEEE International Conference on Granular Computing (GRC 2007), 2007, pp. 708–708. doi:10.1109/GrC.2007.40.

[16] L. M. de Campos, A. Cano, J. G. Castellano, S. Moral, Bayesian networks classifiers for gene-expression data, in: 11th International Conference on Intelligent Systems Design and Applications, 2011, pp. 1200–1206. doi:10.1109/ISDA.2011.6121822.

[17] B. Zhao, M. Yang, H. Diao, B. An, Y. Zhao, Y. Zhang, A novel approach to transformer fault diagnosis using idm and naive credal classifier, International Journal of Electrical Power & Energy Systems 105 (2019) 846 – 855. doi:10.1016/j.ijepes.2018.09.029.

[18] M. Zaffalon, K. Wesnes, O. Petrini, Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data, Artificial Intelligence in Medicine 29 (1) (2003) 61 – 79, artificial Intelligence in Medicine Europe AIME '01. doi:10.1016/S0933-3657(03)00046-0.

[19] R. Bhuvaneswari, K. Kalaiselvi, Naive bayesian classification approach in healthcare applications, International Journal of Computer Science and Telecommunications 3 (1) (2012) 106–112.

[20] B. Cestnik, Estimating probabilities: A crucial task in machine learning, in: Proceedings of the 9th European Conference on Artificial Intelligence (ECAI'90), London Pitman Publishing, 1990, pp. 147–149.

[21] B. Cestnik, I. Bratko, On estimating probabilities in tree pruning, in: European Working Session on Learning (EWSL-91), Vol. 482 of Lecture Notes in Computer Science (LNCS), Springer Berlin Heidelberg, 1991, pp. 138–150. doi:10.1007/BFb0017010.

[22] J. Abellán, Uncertainty measures on probability intervals from the imprecise dirichlet model, International Journal of General Systems 35 (5) (2006) 509–528. doi:10.1080/03081070600687643.

[23] J. Abellán, Equivalence relations among dominance concepts on probability intervals and general credal sets, International Journal of General Systems 41 (2) (2012) 109–122. doi:10.1080/03081079.2011.607449.

[24] M. Lichman, UCI machine learning repository (2013).
URL http://archive.ics.uci.edu/ml

[25] I. H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition, Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.

[26] U. Fayyad, K. Irani, Multi-valued interval discretization of continuous-valued attributes for classification learning, in: Proceeding of the 13th International joint Conference on Artificial Inteligence, Morgan Kaufmann, 1993, pp. 1022–1027.

[27] J. Demšar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research 7 (2006) 1–30.

[28] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, The Annals of Mathematical Statistics 11 (1) (1940) 86–92. doi:10.1214/aoms/1177731944.

[29] P. Nemenyi, Distribution-free multiple comparisons, Doctoral dissertation, Princeton University, New Jersey, USA (1963).