Decision-Making of Homogeneous Multiple Classifiers Based on Attribute Characterisation by Discretisation

Urszula Stańczyk[1] [0000−0002−5071−7187], Beata Zielosko[2] [0000−0003−3788−1094], and Grzegorz Baron[1] [0000−0001−8613−631X]

[1] Department of Computer Graphics, Vision and Digital Systems, Silesian University of Technology, Akademicka 2A, 44-100 Gliwice, Poland
{urszula.stanczyk,grzegorz.baron}@polsl.pl

[2] Institute of Computer Science, University of Silesia in Katowice, Będzińska 39, 41-200 Sosnowiec, Poland
beata.zielosko@us.edu.pl

Abstract. The paper presents research focused on the decision-making process of multiple classifiers, conditioned by the characterisation of attributes provided by supervised discretisation. This transformation of the input domain imposes a specific distribution of data and features, exploited by the homogeneous ensembles of estimators based on the informativeness of attribute domains in a dataset. The committees of inducers aggregated decisions through several defined voting scenarios. The procedure was applied to two classifiers that worked on selected publicly available datasets with different properties. Performance was studied with particular attention given to characteristics and irregularities of input domains before and after discretisation, sensitivity of learners to various data forms, and consequences of the employed voting schema.

Keywords: Voting · Aggregating Decisions · Multiple Classifier · Discretisation · Decision-Making.

1 Introduction

Decision-making is especially important in the field of machine learning or expert systems [23]. Among many influential factors, the level of complexity of such processes depends not only on the number, but also on the types of criteria, the nature, and characteristics of the input domain. Any changes in these elements have an impact on the effectiveness of data exploration in knowledge discovery issues, resulting in variations in the performance of learning systems.

The decision on the assignment of an object to a class can be reached based on one model or through the collaborative work of several approaches. Exploiting their properties, mode of operation, and specificity leads to the construction of multiple models [17]. Classifier ensembles take advantage of model diversity

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

to improve predictive accuracy, reduce errors, and increase the stability of results [19]. Models created by the same learning algorithm are called homogeneous. With multiple suggested decisions, some voting strategy is needed to reach a final verdict, such as majority or weighted voting [4].

Constant technological progress brings increasing volumes of data that vary in their structures, the perspective with which they describe the studied problem, and their physical location. It necessitates the development of techniques that enable the processing of diverse and dispersed data. Some data unification can be needed to simplify decision aggregation [1]. The appropriate preparation of the data, especially in the context of its adaptation to the learning algorithm, constitutes an important step in the data mining process. The outcome of this step is translated into the decisions proposed by the model.

Discretisation is one of the methods used in the preparation of data for analysis. It transforms the values of continuous attributes into a finite number of intervals (called bins) [12]. Therefore, to some extent, it always changes the characteristics of the data and has some influence on the predictions. The transformation can be used to simplify the data or prepare it for learners which require categorical values. The process can be performed using different approaches. In unsupervised discretisation methods, information about the values of the decision attribute is omitted during the process of attribute domain transformation. Supervised discretisation algorithms construct intervals that are most supportive to class distinction. This property can be used to evaluate features, causing their reduction when a single bin is assigned to represent their entire domains.

In the research works presented, the characterisation of attributes provided by a supervised discretisation procedure was exploited to disperse data and attributes. Different groups of features, represented either in the continuous of discrete domains, were allotted votes, used through the several voting scenarios. These were implemented by homogeneous multiple classifiers based on the two popular machine learning algorithms, Naive Bayes and PART, implemented in WEKA workbench [24]. The two inducers operated on several datasets with varying characteristics. Performance was evaluated by labelling samples from test sets, discretised in dependent and independent modes.

The wide ranges of results obtained, observed for various methods of aggregating decisions, point to relations between data formats and transformations of the input domain and the sensitivities of inducers used in the research. The enhanced predictions cannot always be guaranteed, as they are conditioned by many factors. However, the presented mode of operation of a voting classifier can offer conditions beneficial to performance. With such findings, the experiments validated the investigated approach to the creation of classifier committees based on the different characteristics of the attribute domains in a dataset.

The paper consists of five sections. Section 2 is devoted to the approaches of ensemble learning and popular voting techniques. Section 3 includes descriptions of applied discretisation approaches and modes, learning algorithms, and used voting scenarios. Observations on the results of the experiments performed are presented in Section 4. Conclusion and future plans are given in Section 5.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

2 Ensembles and Voting as Means of Reaching a Decision

In the field of decision-making a single model, also known as an underlying or base model, may not work well individually. However, when weak models are aggregated to cooperate through a committee or ensemble [1], they can form a strong model. The idea is based on the concept of 'wisdom of the crowd', which suggests that the decision-making of a larger group of people is typically better than that of an individual, even when this individual is an expert.

An ensemble of classifiers (multiple classifier) trained using the same learning algorithm is referred to as homogeneous, whereas an ensemble generated from estimators trained using different methods is called heterogeneous [15]. The collaboration of multiple base inducers allows different perspectives to be taken into account and negates the impact of individual uncertainties [2]. It reduces variance or deviation, errors, providing better stability and robustness of models, and increasing the accuracy of predictions for the collaborative system.

Combining multiple judgments raises the issue of not only hearing their individual voices, but also establishing some means of reaching the final decision. To this end, voting can be implemented [9]. A classifier committee can employ majority voting, weighted voting, or soft voting. Majority voting exists when each model votes for one class, and the one that receives the most votes is selected. With weighted voting, each model is assigned a weight, and the result depends on the sum of the votes multiplied by the model weights. The soft voting method applies when the models return the probabilities for each class. The final decision is made based on the average probability for each class.

In the context of ensemble learning methods [25], bagging, boosting, and stacking should be mentioned. Bagging involves creating different subsets of samples of the training data, which are randomly selected with replacement from the original dataset. Each subset of the training data is used to train a different classifier of the same type. The individual estimators are then combined by majority vote on their decisions. For each instance, the class selected by the majority of the classifiers is the decision of the ensemble. A popular algorithm using the bagging technique is Random Forests, built from decision trees [6]. Some training parameters of individual trees can be changed randomly, for example, bootstrapped replicas of the training data, or feature subsets.

Boosting involves creating a series of simple inducers, where each new model focuses on correcting errors made by previous ones. The process is sequential and each model builds on the results of the previous one. The aim of boosting is to create a strong estimator by combining the results of several weak classifiers, ultimately improving the overall accuracy of the model. The final decision is made based on a weighted vote or aggregation of the results of all weak classifiers.

Stacking improves prediction accuracy by integrating the results from different base learners [17]. The approach consists of multiple base classifiers and a meta-classifier. Each base estimator is trained separately using a distinct learning algorithm to perform the classification task, and their combined results are used by the meta-classifier to generate the final prediction.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

3 Framework of Experiments

In the research several datasets with varied characteristics were used, always with a noticeable proportion of attributes that after supervised discretisation were assigned single intervals. The characterising property of this process was used to distribute data. Based on continuous and discrete forms of features, an approach to aggregating decisions was defined that involved several voting scenarios. The procedure was applied for homogeneous multiple classifiers whose performance was studied. The section details the research stages.

3.1 Datasets and Attribute Domains

To provide a wide scope for observations, in the research four pairs of datasets were used. Three pairs used data available in the UCI Machine Learning Repository [13]. These were, respectively, Avila [8], Magic [5] and Wave [7]. The fourth pair, Style, relied on data from the stylometric domain [18]. Within each pair, the datasets were simply enumerated 1 and 2, resulting in Avila1 and Avila2, Magic1 and Magic2, Wave1 and Wave2, Style1 and Style2.

The datasets can be grouped into two categories depending on the number of attributes. Wave and Style contained roughly twice as many characteristic features as Avila and Magic. There were no missing values. The single decision attribute was categorical, while the condition attributes (criteria) had continuous values. Each of the eight datasets consisted of a single train and two test sets. Each set was prepared for a binary classification task with balanced data. Both classes were assumed to be of the same importance, with the same misclassification costs, so the performance of the predictors was evaluated by classification accuracy [22] obtained for the test sets, over which the average was calculated.

3.2 Data Transformation and Distribution

In the research two popular discretisation approaches were used: the supervised Kononenko algorithm (denoted dsK) [14], and unsupervised method of equal width binning (denoted duw) [10]. Supervised approaches take into account specific attribute values, but also how distribution of datapoints in the input domain relates to information on classes, while unsupervised procedures focus entirely on the domain of the transformed attribute.

The Kononenko algorithm is based on the calculation of the conditional entropy. Taking into account the binary discretisation of a continuous attribute A with respect to set S, candidate cut-points T of the discretised attribute are tested referring to the changes in entropy caused by them. For the optimal cut-point T_{opt} that splits the set S into two subsets, class information entropy $E(A, T_{opt}; S)$ is minimal, and this cut-point is selected. This process is applied recursively until the inequality (1) is satisfied:

$$\log \binom{N}{N_{C_1} \dots N_{C_k}} + \log \binom{N+k-1}{k-1} >$$

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

$$\sum_j \log \binom{N_{A_j}}{N_{C_1 A_j} \dots N_{C_k A_j}} + \sum_j \binom{N_{A_j} + k - 1}{k - 1} + \log N_T, \tag{1}$$

where N is the number of training instances, and N_{C_i} the number of training instances from the class C_i. N_{A_x} denotes the number of instances with x-th value of the given attribute, $N_{C_i A_y}$ the number of instances from class C_i with y-th value of the given attribute. N_T is the number of possible cut-points. The method works in a top-down fashion. It starts with one interval, divided into subintervals until a stopping condition is reached. Therefore, it is possible that for some variable a single interval is assigned as a categorical representation.

The equal width binning method is considered simple. It consists of the following steps: (i) sorting the values of a continuous attribute, (ii) designating the minimum and maximum values of the processed attribute, and (iii) dividing the range into the k equal width discrete intervals, where k is a parameter defined by a user. The advantage of this method is speed. It is also directly controlled by the value of the parameter k, which can result in an uneven distribution of values if the datapoints are clustered in a certain range.

Supervised discretisation can be considered a mechanism for feature reduction. If there are some variables to which a single bin is assigned to represent their entire domains, then in a discrete domain these attributes have zero informative content and can be discarded. However, if processed by some other approach (for example, an unsupervised discretisation algorithm), they can still be found useful. In the research, the characterisation of attributes by Kononenko discretisation (shown in Table 1) was employed as a controlling factor when the features in the datasets were distributed into two categories in order to separate 1 bin variables from multi-bin variables.

Table 1. Attributes in datasets and characteristics found by supervised discretisation of train sets with the Kononenko algorithm.

Number of attributes		Dataset							
		Avila1	Avila2	Magic1	Magic2	Wave1	Wave2	Style1	Style2
Total		10	10	10	10	21	21	20	20
Number of bins	1	5	4	3	4	8	8	8	10
	2	3	4	6	6	10	7	7	8
	3	1	1	1		3	5	4	2
	4	1	1				1	1	

To some extent, these numbers of discrete bins assigned to attributes can be treated as measures of importance for characteristic features [20], showing the complexity of the relations between attribute values and the distinction of classes. The minimal value of 1 means that a variable is estimated as irrelevant, and for each of the used datasets the number of such features was at least 30% and at most 50%. The higher numbers of bins indicate that specific discrete categorical representations need to be distinguished to support recognition of classes. However, in the input continuous domain all variables have non-zero informative content. Rejecting any of them has an impact on classifier performance. Whether

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

it works to advantage or disadvantage depends on the data but also on the mode of operation of the inducer and its sensitivity to changes in data format.

When a discretisation process is applied to datasets that are composed of several sets with the same features (such as a train and test sets), translation of attribute domains can be performed in several ways [3]. The simplest to execute is the independent transformation, where the domain characteristics are studied only locally, within each set separately, and regardless of all other sets. This approach can lead to establishing for a variable not only different cut-points between intervals, but also different numbers of bins, in particular for supervised discretisation algorithms. In a way, it can be treated as constructing discrete data models, the comparison of which then becomes a part of pattern recognition and classification. Another way of proceeding involves imposing definitions of intervals built for one set (typically a train set) on the other set (a test set). This processing path completely disregards the local characteristics of the datapoints in the test sets.

Depending on the irregularities present in the data [21], both types of discretisation procedures applied to datasets can be beneficial or detrimental to the evaluation of performance. Therefore, both were used in the research. For each studied dataset, when its parts were discretised, the two test sets included were subjected to independent processing of all sets, and then they were denoted as test independent, Tind. When discrete test sets were constructed by application of definitions for bins formed for the corresponding train sets, they were denoted as test on learn, ToL.

3.3 Data Mining Approaches Used

In the research works, for the construction of homogenous multiple classifiers, the Naive Bayes (NB) and PART algorithms were used, implemented in WEKA software [24]. Both classifiers are capable of operating on numeric and categorical variables, which was a requirement from the perspective of intended experiments. They have a different mathematical background and exhibit different sensitivity when data is transformed by discretisation.

The Naive Bayes classifier is often used as a reference model. It is based on Bayes' theorem, which assumes that all attributes are independent of each other. Based on the features that describe the given object, it calculates the probability of an object belonging to particular classes and then selects the class label with the highest probability [16]. It is widely used in machine learning applications, for example, natural language processing, spam filtering, and sentiment analysis.

PART [11] algorithm relates to decision rules and decision trees. It generates partially constructed C4.5 decision trees and derives a rule from each one. The rules are induced within the framework of the separate-and-conquer approach. When a rule is constructed, all instances covered by this rule are repeatedly removed until all instances are covered. The rule construction stage differs from standard separate-and-conquer methods in that a partially pruned decision tree is constructed for a set of instances. The leaf to which the largest number of

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

objects is assigned is transformed into a rule, and then the tree is discarded. At the end of the training stage, a list of decision rules is obtained.

3.4 Voting Scenarios for Homogenous Multiple Classifiers

The two selected machine learning algorithms were exploited in the research to obtain a homogenous multiple classifier, that is, one type of learner working on diversified data. The diversification and distribution of the data was controlled by characteristics of the input space discovered by supervised discretisation. The data preparation procedure is shown in Fig. 1. To reach a final decision, the complex classifiers employed simple majority voting, through several defined voting schemas.

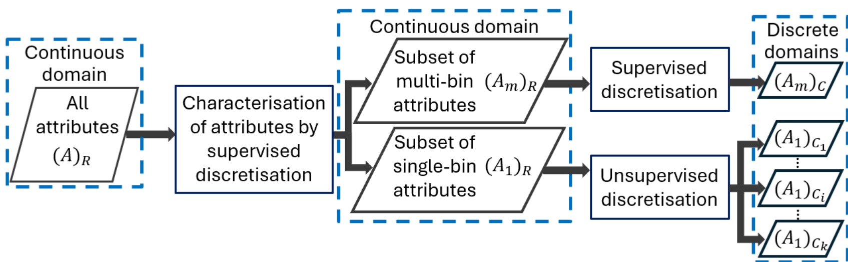

Fig. 1. Procedure of data preparation for multiple classifiers using voting scenarios.

In all voting scenarios, three votes were considered. The individual votes came from data distributed with the help of characterisation property of supervised discretisation procedure applied to attributes that caused assigning multiple or single intervals to represent discrete attribute domains. The votes were obtained either in a single step or in two, causing one level and two level voting. The prevailing influence was given either to multiple-bin variables or to 1-bin features. The votes were always based on knowledge learnt from some combination of continuous and discrete data.

Let A denote the entire set of features. $(A)_R$ denotes representation in the continuous domain, while $(A)_C$ stands for a discrete representation. The set A is partitioned into two subsets, A_m and A_1 ($A = A_m \cup A_1$), the former including such variables that were assigned multiple intervals by supervised discretisation, and the latter the features with just single bins. Then $(A_m)_R$ denotes the group of multi-bin variables and $(A_1)_R$ the set of single-bin features, both represented in the continuous domain. $(A_m)_C$ is used to indicate the set of multi-bin variables represented in the discrete domain obtained by the supervised Kononenko procedure. $(A_1)_C$ denotes the set of single-bin variables in a discrete domain resulting from unsupervised equal width binning with the varying number of intervals, so for $(A_1)_{C_i}$, i defines the number of bins required.

The voting scenarios (VS) (visualised in Fig. 2), explored through the experiments conducted on the datasets, were as follows:

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

- VS1 denoted $(A_m)_R(A_m)_C(A_1)_R$—single level voting, multi-bin variables carry the majority as they are given two votes over the one of single-bin variables, both groups of features considered in the continuous domain, but with added impact of characterisation by supervised discretisation. This scenario of aggregating decisions can be seen as taking advantage of supervised discretisation. It sifts through variables and selects those that are important, yet with keeping those found unimportant still in considerations and their original informative content in the continuous domain. Thus two votes rely on the continuous form of variables and one on the discrete,
- VS2 denoted $(A_m)_R(A_m)_C(A_1)_C$—differs from the first in forms of single bin variables, as they are employed in their discrete forms in the process of decision making. This is possible by the application of some unsupervised discretisation algorithm, such as equal width binning. Since the procedure requires the input parameter i specifying the number of bins to be constructed, several variants $(A_m)_R(A_m)_C(A_1)_{C_i}$ are obtained, and they are treated separately. Therefore, two votes are based on discrete representation of data and one on continuous,
- VS3 denoted $(A_m)_R(A_m)_C\,((A_1)_{C_1}\ldots(A_1)_{C_k})$—similar to the second voting scenario as refers to the same forms of variables, but with two level voting. On one level, all k constructed data variants obtained for single-bin variables by unsupervised discretisation agree on a decision by simple majority voting. This decision is passed on to the second level with votes from multi-bin variables in their continuous and discrete forms. Still two votes rely on discretised data, but in this scenario the influence of a particular data variant, indicated by the input parameter i of unsupervised discretisation, is minimised,
- VS4 denoted $(A_m)_R(A_1)_R\,((A_1)_{C_1}\ldots(A_1)_{C_k})$—two level voting with the prevailing votes of single bin variables, which are considered in both continuous and discrete forms. On the other hand, two votes come from continuous attributes distributed between the two categories, which is supported with the third vote aggregated through voting of all discrete variants of 1-bin features transformed by unsupervised algorithm,
- VS5 denoted $(A_m)_C(A_1)_R\,((A_1)_{C_1}\ldots(A_1)_{C_k})$—two level voting similar to the forth voting schema, but with using only discrete forms of multi-bin variables. With this approach, two votes rely on discrete data, taking into consideration all k unsupervised variants, which is supported by the original continuous form of 1-bin features.

4 Observations on Performance for Voting Classifiers

Depending on their own specific sensitivity to the input data form, domain transformation by discretisation procedures can cause inducers to react in various ways. Some can be simply enabled, if they can operate only on nominal or discrete features. For others the translation can be beneficial, while still other classifiers can be harmed by reduction of accessible information reaching too far.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

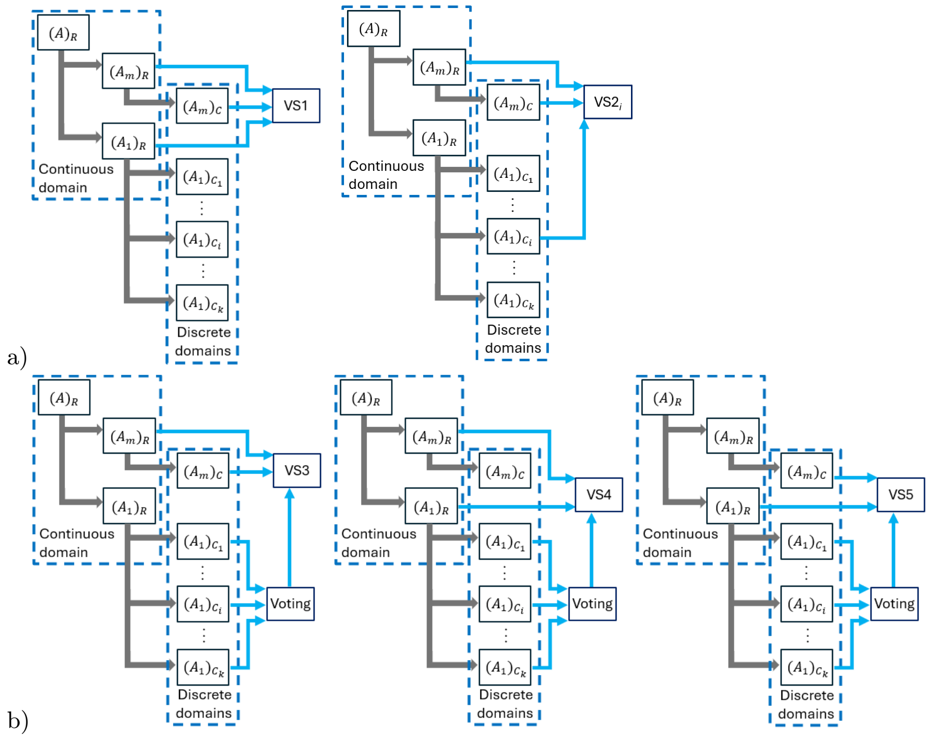

Fig. 2. Data, domains and voting scenarios investigated, with voting in: a) single level, b) two levels.

If learners possess their own inherent mechanism that corresponds to discretisation procedure, then any consequences of such transformation of the input domain can become obscured, indistinguishable from other processing.

The Naive Bayes and PART classifiers that were used in the research can efficiently work both in the continuous and discrete domain. The performance detected in the continuous domain is given in Table 2. For the Avila and Magic datasets, operating on smaller numbers of attributes, PART outperformed the Naive Bayes classifier. The opposite can be stated for Wave and Style datasets, where Naive Bayes was better than PART, however, the differences were less noticeable.

Table 2. Performance [%] of inducers operating on the datasets in the continuous domain $(A)_R$.

Inducer	Dataset							
	Avila1	Avila2	Magic1	Magic2	Wave1	Wave2	Style1	Style2
NB	61.04	73.75	67.78	71.11	88.13	88.26	98.75	85.07
PART	81.74	91.88	71.94	72.08	76.18	83.68	97.08	79.10

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

Since discretisation by the Kononenko algorithm was a controlling factor for data distribution, it was reasonable to test the employed inducers on the data subjected to this transformation as the only processing step. The observed performance, given in Table 3, is provided for the labelling of samples from the test sets that were discretised independently on train data (denoted Tind), and for test sets transformed based on intervals constructed for train data (denoted ToL). Generally, discretisation caused mixed results, in some cases improved performance was detected while in others some decrease occurred.

Table 3. Performance [%] of inducers operating on the datasets in the discrete domain obtained by the Kononenko algorithm $(A)_C$.

		Dataset							
Inducer	Test set	Avila1	Avila2	Magic1	Magic2	Wave1	Wave2	Style1	Style2
NB	Tind	63.89	50.00	50.00	50.00	95.35	88.75	92.15	64.38
	ToL	90.07	93.54	64.24	70.76	86.94	88.19	98.19	78.47
PART	Tind	72.78	10.63	77.85	50.00	88.68	89.79	92.08	65.49
	ToL	90.63	87.64	77.22	70.21	83.06	85.90	95.90	81.46

For the most part, independent discretisation of test sets brought worse levels of correct predictions than when relying on interval definitions learnt for train sets. It shows data irregularities observable in sets that refer to the same variables, but are considered separately. The exception to this trend was visible for both Wave datasets for both inducers, and for PART for Magic1. The marked differences in performance for both classifiers working on continuous versus discrete data gave reasons to expect variations in accuracy for decision-making processes conditioned by data form, as defined by all investigated voting scenarios.

For the first voting schema, VS1, with predictions shown in Table 4, three votes are considered in one level of aggregating decisions. The multi-bin variables have two votes (one based on continuous domain and the other on discrete) against a single vote given to 1-bin features, which means that the latter have influence on the decision-making process only when continuous and discrete forms of multi-bin attributes differ in their proposed labels for a test sample.

Table 4. Performance [%] of inducers operating with (VS1) Voting Scenario: $(A_m)_R(A_m)_C(A_1)_R$.

		Dataset							
Inducer	Test set	Avila1	Avila2	Magic1	Magic2	Wave1	Wave2	Style1	Style2
NB	Tind	65.69	74.38	64.31	68.82	92.43	87.64	97.71	76.67
	ToL	78.82	82.64	66.04	72.01	88.13	88.26	98.75	82.78
PART	Tind	82.22	60.63	77.85	72.01	81.25	83.47	93.33	70.42
	ToL	90.56	90.63	77.22	73.75	78.96	86.11	94.65	82.64

When independently processed test sets are considered, this voting scenario gave better results than when working in dsK discrete domain (for $(A)_C$) for the Avila, Magic and Style datasets, while for Wave the accuracy was degraded for

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

both classifiers. For ToL test sets, Naive Bayes returned higher predictions for all datasets apart from Avila, while for PART the accuracy in four cases slightly increased, in three decreased, and once was the same.

Comparison of this voting scenario with the performance for the data before any transformations or distribution leads to the conclusion that for NB classifier only for the Avila datasets clear advantages can be observed regardless of the test set type. For other datasets, accuracy was varied and not necessarily maintained. PART indicated different trends, showing its own sensitivity to the processing path, and improved results for Avila1, Magic1, Wave1 for both types of test sets and for Magic2, Wave2, Style2 for ToL test sets only.

The second voting scenario (VS2) differs from VS1 in the form of one component, the A_1 features being considered in discrete domains. It involved the construction of several variants of data, because attributes that were assigned single bins by supervised discretisation were transformed with an unsupervised procedure. The number of intervals varied from two to ten. The performance reported by the inducers is provided in Table 5. For each dataset, each learner and the type of test set used in evaluation, the bolded entries indicate detected maximum. An analysis immediately calls attention to the high diversity of results obtained. However, both inducers show improvements over the first voting scenario for both types of test sets for the majority of datasets.

The performance observed for multiple classifiers operating according to the third voting scenario (VS3) is presented in Table 6. This schema relies on the same representations for groups of attributes as in VS2, but with a very noticeable difference: the decision-making process is executed in two levels. The internal level involves reaching a decision by taking the votes from all discrete data variants obtained by unsupervised transformations. Once this decision is available, it is passed on to the second level of voting, where again the attributes with multiple bins defined by supervised discretisation play the leading role.

For both types of test sets, Naive Bayes performed mostly better, with the exception of the Wave1 and Style1 datasets, for which the accuracy was slightly decreased with respect to the best cases observed when VS2 was employed. PART for all datasets exhibited decreased performance. Clearly in this case, aggregating decisions between data variants resulting from unsupervised discretisation led to the best cases being over-voted by others, less advantageous to predictions.

For Voting Scenario 4 (VS4), the results are shown in Table 7. It employs two levels of voting and starts in the same way as VS3, making a decision based on votes referring to the variables (A_1) represented in all discrete variants considered after unsupervised discretisation with equal width binning. In this case, at the second level of aggregating decisions, $(A_m)_C$ is replaced by $(A_1)_R$. So, at this level, two votes are based on 1-bin variables and only one on A_m features.

In consequence of these changes with respect to VS3 (and VS1 and VS2), for both inducers, for most datasets, performance decreased. The exception can be observed for independently discretised test sets and Avila2 and Magic1 for NB, and Avila2 and Magic2 for PART, where some increase can be noted in comparison to the third voting scenario. The suggestions of the attributes A_1,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

Table 5. Performance [%] of inducers operating with (VS2) Voting Scenario: $(A_m)_R(A_m)_C(A_1)_C$.

Inducer	Data variant	Test set	Dataset							
			Avila1	Avila2	Magic1	Magic2	Wave1	Wave2	Style1	Style2
NB	uw02	Tind	65.14	62.22	55.56	68.06	**94.65**	**88.75**	96.60	71.94
		ToL	**86.60**	84.51	**66.60**	70.42	**88.75**	88.26	98.82	81.60
	uw03	Tind	**71.46**	53.26	**57.36**	66.18	91.81	87.64	97.71	74.58
		ToL	76.25	84.38	66.04	72.64	87.57	88.26	98.75	82.15
	uw04	Tind	68.68	51.74	55.56	66.11	93.54	88.19	97.15	73.33
		ToL	76.67	84.58	66.04	71.46	87.57	88.26	98.82	82.71
	uw05	Tind	71.04	**66.39**	55.56	65.21	92.99	87.01	97.71	72.71
		ToL	72.01	83.89	65.49	**73.26**	86.94	88.26	98.75	**82.78**
	uw06	Tind	70.42	53.89	54.93	**69.24**	91.81	88.26	98.26	74.03
		ToL	74.44	83.33	66.04	72.01	87.57	88.26	98.75	82.22
	uw07	Tind	69.86	50.56	55.00	65.83	92.85	88.26	**98.82**	**77.08**
		ToL	74.93	81.53	65.49	72.01	87.50	88.26	**99.38**	82.22
	uw08	Tind	68.75	62.15	56.81	68.61	92.92	87.01	97.15	72.78
		ToL	73.13	84.44	66.04	72.01	87.01	88.26	98.75	82.22
	uw09	Tind	69.24	61.39	56.11	66.39	93.54	88.26	98.26	**77.08**
		ToL	72.57	82.64	66.04	71.46	87.01	88.26	98.75	**82.78**
	uw10	Tind	69.86	59.51	56.67	66.46	93.47	88.26	97.71	73.96
		ToL	74.38	**85.69**	**66.60**	71.46	87.57	88.26	98.75	**82.78**
PART	uw02	Tind	79.17	53.54	73.06	66.18	83.61	85.35	94.51	77.08
		ToL	89.93	90.07	74.24	71.94	82.43	84.38	97.50	83.82
	uw03	Tind	84.24	51.04	79.10	51.67	84.24	91.11	94.51	75.76
		ToL	91.74	89.51	**76.60**	70.07	82.57	86.53	96.88	81.60
	uw04	Tind	88.19	53.06	76.81	63.96	83.96	87.15	94.38	77.22
		ToL	90.56	90.63	73.13	70.90	80.21	86.60	96.94	83.89
	uw05	Tind	79.17	54.17	**79.17**	63.26	81.81	89.93	**95.76**	76.39
		ToL	90.00	90.07	75.49	70.28	80.14	85.28	**98.26**	**84.44**
	uw06	Tind	78.19	53.13	75.97	56.04	82.43	86.25	93.89	77.15
		ToL	91.25	89.44	**76.60**	69.17	81.32	86.67	95.14	83.82
	uw07	Tind	88.40	49.44	71.39	64.38	85.97	85.21	91.46	72.08
		ToL	**92.36**	**91.25**	71.94	71.32	**84.79**	**87.85**	96.94	82.85
	uw08	Tind	89.51	47.78	73.54	**68.68**	82.43	90.07	95.14	75.83
		ToL	91.81	88.96	75.97	**77.43**	80.07	85.35	98.13	82.64
	uw09	Tind	**90.69**	47.08	71.39	55.14	**87.01**	**91.25**	92.08	72.22
		ToL	91.18	90.07	71.94	71.39	82.43	84.79	93.54	84.31
	uw10	Tind	77.36	**59.65**	77.22	64.44	80.63	84.65	93.75	**78.33**
		ToL	90.56	90.69	**76.60**	69.65	80.21	84.93	95.21	81.53

which were deciding in this case, turned out to be not sufficiently accurate to arrive at higher levels of correct predictions. This confirmed the estimation of their poor relevance for the task obtained by supervised discretisation.

The fifth voting schema, VS5, is similar to the VS4, with one difference: variables that by supervised discretisation received multiple intervals are used in their discrete form. A_1 features are employed in both continuous and discrete forms, and from the latter in one level of voting, the decision is established before it is passed on to the second level of aggregating decisions. The results of such processing are included in Table 8. It can be observed that the change in the domain from continuous to discrete for the A_m variables caused relatively small differences with respect to performance for VS4. These differences go both ways, for some datasets, to advantage, while for others, to disadvantage. For Naive Bayes and Tind test sets, decreased performance was detected more often, but for ToL test sets, improvement was more frequent. For the PART algorithm for both types of test sets, worsened results were noted in almost all cases.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

Table 6. Performance [%] of inducers operating with (VS3) Voting Scenario: $(A_m)_R(A_m)_C\,((A_1)_C \ldots (A_1)_C)$.

Inducer	Test set	Dataset							
		Avila1	Avila2	Magic1	Magic2	Wave1	Wave2	Style1	Style2
NB	Tind	70.49	54.51	55.00	68.13	93.54	88.26	97.71	74.58
	ToL	72.64	83.82	66.04	72.01	87.57	88.26	98.75	82.78
PART	Tind	84.65	53.33	74.17	60.35	84.10	87.50	93.82	77.64
	ToL	90.00	90.07	75.97	68.47	81.88	84.79	96.25	82.78

Table 7. Performance [%] of inducers operating with (VS4) Voting Scenario: $(A_m)_R(A_1)_R\,((A_1)_C \ldots (A_1)_C)$.

Inducer	Test set	Dataset							
		Avila1	Avila2	Magic1	Magic2	Wave1	Wave2	Style1	Style2
NB	Tind	62.01	67.22	63.19	59.65	58.13	59.65	77.15	68.47
	ToL	59.10	59.10	64.31	57.85	55.07	56.60	74.93	66.39
PART	Tind	72.29	78.68	64.79	65.49	63.68	73.96	70.90	75.97
	ToL	66.39	78.13	66.53	64.79	67.71	71.25	72.64	76.46

The wide ranges of results observed for various methods of aggregating decisions show relations between data formats and transformations of the input domain and the sensitivities of inducers used in the research. The enhanced predictions cannot be guaranteed, as the outcome is conditioned by many factors. However, depending on characteristics of the attribute domains, local properties of each processed set, and approach used to discretisation of test sets employed in evaluation of performance, the presented mode of operation of a voting classifier can benefit from dispersing data and taking into account characterisation of features by supervised discretisation. When those attributes which are thus found as supporting distinction of classes play the leading role in decision-making but are also supported with suggestions from 1-bin variables, the more involved process of pattern recognition and classification can lead to improved accuracy.

5 Conclusions

Voting constitutes one of the collaborative approaches to a decision-making process involving multiple criteria and classification committees. To reach a final decision on assigning a class label to a studied object, suggestions are taken from a number of models that represent the knowledge discovered by data mining. In the research described in this paper the models were learnt from data dispersed based on the characterising property of supervised discretisation. Through the assigned intervals, the transformation evaluates the complexity of the relations between the attributes and their values and the distinction of classes, which was exploited for dividing the features into categories. Votes allotted to different groups and the forms of variables were next used within the several investigated one-level and two-level voting scenarios for selected homogeneous multiple learning systems, capable of operating in both continuous and discrete domains. The performance was evaluated by labelling test sets discretised in two distinc-

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

Table 8. Performance [%] of inducers operating with (VS5) Voting Scenario: $(A_m)_C(A_1)_R\,((A_1)_C\ldots(A_1)_C)$.

Inducer	Test set	Dataset							
		Avila1	Avila2	Magic1	Magic2	Wave1	Wave2	Style1	Style2
NB	Tind	62.22	48.61	55.00	58.96	59.24	59.03	76.04	64.17
	ToL	71.39	62.64	64.31	57.85	54.51	56.60	74.93	66.39
PART	Tind	67.64	39.72	68.47	54.93	67.64	78.96	70.28	68.75
	ToL	65.83	76.18	67.78	59.51	67.01	69.93	72.15	76.46

tively different ways, allowing to investigate irregularities present in the several datasets explored.

The approach to classifier ensemble construction based on attribute domain characteristics was investigated for five defined scenarios, including majority and aggregation voting techniques with Naive Bayes and PART algorithms. The results from the experiments showed conditions for possible improved predictions and the effects of dependent and independent discretisation modes on the performance of the classifier. They confirmed the validity of the research framework.

In future research, the application of other inducers with different modes of operation will be studied. Furthermore, other schemas for aggregation decisions by voting will be defined, involving combinations of learners dependent on their sensitivity to data form. Heterogeneous ensembles will also be researched.

Acknowledgments. The research works presented in the paper were carried out within the statutory project of the Department of Computer Graphics, Vision and Digital Systems (RAU-6, 2025), at the Silesian University of Technology, Gliwice, Poland, and at the Institute of Computer Science, University of Silesia in Katowice, Sosnowiec, Poland.

References

1. Azaiez, M.N.: Aggregation in decision problems: Concepts and applications. Journal of King Saud University - Engineering Sciences **21**(2), 101–109 (2009)
2. Baron, G.: Analysis of multiple classifiers performance for discretized data in authorship attribution. In: Czarnowski, I., Howlett, J.R., Jain, C.L. (eds.) Intelligent Decision Technologies 2017: Proceedings of the 9th KES International Conference on Intelligent Decision Technologies – Part II, Smart Innovation, Systems and Technologies, vol. 72, pp. 33–42. Springer International Publishing (2018)
3. Baron, G., Harężlak, K.: On approaches to discretization of datasets used for evaluation of decision systems. In: Czarnowski, I., Caballero, A.M., Howlett, R.J., Jain, L.C. (eds.) Intelligent Decision Technologies 2016, pp. 149–159. Springer International Publishing, Cham (2016)
4. Bauer, E., Kohavi, R.: An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. Machine Learning **36**(1), 105–139 (1999)
5. Bock, R.: MAGIC Gamma Telescope (Dataset). UCI Machine Learning Repository (2004), last accessed 2025/02/05
6. Breiman, L.: Random forests. Machine Learning **45**, 5–32 (2001)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17

7. Breiman, L., Stone, C.: Waveform Database Generator (Version 1). UCI Machine Learning Repository (1984), last accessed 2025/02/05
8. De Stefano, C., Maniaci, M., Fontanella, F., Scotto di Freca, A.: Reliable writer identification in medieval manuscripts through page layout features: The "Avila" Bible case. Engineering Applications of Artificial Intelligence **72**, 99–110 (2018)
9. Dogan, A., Birant, D.: A weighted majority voting ensemble approach for classification. In: 2019 4th International Conference on Computer Science and Engineering (UBMK). pp. 1–6 (2019)
10. Dougherty, J., Kohavi, R., Sahami, M.: Supervised and unsupervised discretization of continuous features. In: Machine Learning Proceedings 1995: Proceedings of the 12th International Conference on Machine Learning. pp. 194–202. Elsevier (1995)
11. Frank, E., Witten, I.H.: Generating accurate rule sets without global optimization. In: Proceedings of the Fifteenth International Conference on Machine Learning. p. 144–151. Morgan Kaufmann Publishers Inc. (1998)
12. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Morgan Kaufmann (2011)
13. Kelly, M., Longjohn, R., Nottingham, K.: The UCI Machine Learning Repository. https://archive.ics.uci.edu, last accessed 2025/02/05
14. Kononenko, I.: On biases in estimating multi-valued attributes. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence IJCAI'95. vol. 2, pp. 1034–1040. Morgan Kaufmann Publishers Inc., Montreal, Canada (1995)
15. Kuncheva, L.I., Whitaker, C.J.: Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Machine Learning **51**(2), 181–207 (2003)
16. Peretz, O., Koren, M., Koren, O.: Naive bayes classifier – an ensemble procedure for recall and precision enrichment. Engineering Applications of Artificial Intelligence **136**, 108972 (2024)
17. Rokach, L.: Ensemble-based classifiers. Artificial Intelligence Review **33**, 1–39 (2010)
18. Rybicki, J., Eder, M., Hoover, D.: Computational stylistics and text analysis. In: Crompton, C., Lane, R., Siemens, R. (eds.) Doing Digital Humanities: Practice, Training, Research, pp. 123–144. Routledge, 1 edn. (2016)
19. Sesmero, M.P., Iglesias, J.A., Magán, E., Ledezma, A., Sanchis, A.: Impact of the learners diversity and combination method on the generation of heterogeneous classifier ensembles. Applied Soft Computing **111**, 107689 (2021)
20. Stańczyk, U.: Evaluating importance for numbers of bins in discretised learning and test sets. In: Czarnowski, I., Howlett, J.R., Jain, C.L. (eds.) Intelligent Decision Technologies 2017: Proceedings of the 9th KES International Conference on Intelligent Decision Technologies – Part II, Smart Innovation, Systems and Technologies, vol. 72, pp. 159–169. Springer International Publishing (2018)
21. Stańczyk, U., Zielosko, B.: Data irregularities in discretisation of test sets used for evaluation of classification systems: A case study on authorship attribution. Bulletin of the Polish Academy of Sciences: Technical Sciences **69**(4), 1–12 (2021)
22. Stąpor, K., Ksieniewicz, P., García, S., Woźniak, M.: How to design the fair experimental classifier evaluation. Applied Soft Computing **104**, 107219 (2021)
23. Taherdoost, H., Madanchian, M.: Multi-criteria decision making (mcdm) methods and concepts. Encyclopedia **3**(1), 77–87 (2023)
24. Witten, I., Frank, E., Hall, M.: Data Mining. Practical Machine Learning Tools and Techniques. Morgan Kaufmann, 3rd edn. (2011)
25. Zhou, Z.H.: Ensemble Methods: Foundations and Algorithms. Chapman & Hall/CRC (2012)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97567-7_17