Int. J. Nonlinear Anal. Appl. 17 (2026) 5, 25–42
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2024.33099.4925

# Improvement accuracy for C4.5 decision tree algorithm

Nima Rasekh, Daniyal Nasiri Bavil*

*Department of Information Engineering, Padua University, Italy*

*(Communicated by Haydar Akca)*

**Abstract**

The decision tree is indeed the most widely used approach to represent classifiers. Initially, it has been studied in the field of decision theory and statistics. However, it was found to be effective in other disciplines, such as data mining, machine learning and pattern recognition. This research deals with the problem of finding the parameter settings of the decision tree algorithm in order to achieve higher accuracy for a given domain. The proposed approach, Improved C4.5 (IC4.5), is a supervised learning model based on the C4.5 algorithm to construct a decision tree. The modification to the C4.5 algorithm includes using improved gain instead of the gain ratio measure to choose the best attribute and increase the accuracy of the decision tree. The introduced algorithm has been experimented with on some data sets from the UCI repository. The results obtained from experiments show that the accuracy of IC4.5 is greater than C4.5 in increasing the accuracy of the decision tree.

Keywords: decision tree, data mining, C4.5 algorithm, IC4.5, accuracy, classifiers
2020 MSC: 05C05, 90C08

## 1 Introduction

Data mining [12] is the process of discovering interesting and useful patterns and relationships in large volumes of data. Due to the large volume of data and the high speed of data production, it is practically impossible to analyze data using traditional methods. Meanwhile, data mining, one of the most popular topics in the present century, has contributed to advancing science and technology in many areas. As a result, researchers have extensively used data mining to analyze data in the recent decade.

With their ability to analyze vast amounts of data and extract patterns, data mining and machine learning techniques can significantly enhance the performance and efficiency of electrical systems, such as in power electronics applications [26, 27, 28, 29, 30, 31, 32, 39, 40]. For example, In the context of wireless power transfer systems [16, 17, 20, 21, 22, 23, 24, 25, 33], machine learning algorithms can be employed to optimize the design and control of the system, enabling efficient power delivery while minimizing losses and maximizing the overall system performance.

Data mining is a process of extracting information and knowledge from a large number of incomplete, noisy, fuzzy and random data. In these data, the information and knowledge are implicit, which people do not know in advance, but are potentially useful. Therefore, constructing fast and accurate classifiers for large data sets is essential in data mining and knowledge discovery. Data Mining is discovering knowledge from vast amounts of data. Data Mining algorithms include classification, clustering, association rules, regression, summarization, and anomaly detection. There are

*Corresponding author
*Email addresses:* n.rasekh1996@gmail.com (Nima Rasekh), Dani.nasiri@gmail.com (Daniyal Nasiri Bavil)

*Received:* January 2024 *Accepted:* March 2024

various proprietary and free open-source data mining tools using which these algorithms can be applied to the given datasets and analyzed.

Data mining methods can be divided into descriptive and predictive analysis according to their realized functions. In the final analysis, descriptive analysis is useful for predictive analysis. It fully reflects the overall distribution of the data and can show the inherent characteristics of the relevant data. Correspondingly, predictive analysis is based on descriptive analysis and treats its analysis results from a developmental perspective, thereby generating a prediction of future data. It gives the final decision-maker a data-level reminder. Predictive analysis mainly refers to prediction based on classification or based on statistical regression problems. At the same time, the main representatives of descriptive analysis are association rule mining and cluster analysis methods. Several standard forms of classification, clustering, and association rules are introduced as follows.

**A. Classification:** Classification is the process of categorizing data into different categories based on a well-defined conceptual description. The naive Bayes method of statistics and the decision tree learning method in machine learning are various implementations of classification techniques. Classification is a data mining technique that maps data into predefined groups or classes. It is a supervised learning method which requires labelled training data to generate rules for classifying test data into pre-determined groups or classes [11]. It is a two-phase process. The first phase is the learning phase, where the training data is analyzed, and classification rules are generated. The next phase is classification, where test data is classified into classes according to the generated rules.

The process of building a classification model typically involves the following steps:

1. Data preparation: This step involves cleaning and pre-processing the data, such as removing missing values and transforming the data into a format that the classification algorithm can use.
2. Model selection: This step involves choosing an appropriate classification algorithm based on the characteristics of the data and the desired outcome. Common algorithms include decision trees, k-nearest neighbors, and support vector machines.
3. Model training: This step involves using the training data to train the classification algorithm and build the model. The model is trained by adjusting its parameters to minimize the difference between predicted and actual class labels.
4. Model evaluation: This step involves evaluating the performance of the classification model on a test dataset that is separate from the training data. This can be done by calculating metrics such as accuracy.
5. Model deployment: This step involves deploying the classification model in a production environment, which can be used to predict new instances.

Some crucial parameters decide which classifiers are suitable for a particular data set, such as follows.

- Accuracy: This includes the classifier's accuracy in predicting the class label. Accuracy can be estimated using one or more test sets. Accuracy is the percentage that a classifier classifies the tuple correctly.
- Speed: How much time is required to construct the model? This also includes the time the model uses to classify the number of tuples (prediction time). In other words, this refers to the computational costs.
- Robustness: This is the ability of the classifier or predictor to make correct predictions given noisy data or data with missing values.
- Scalability: The number of the tuple is independent of the classifiers. Efficiency is calculated in terms of database size.
- Interpretability: Understanding and insight provided by the model. Interpretability is subjective and, therefore, more difficult to assess. Other measures include the goodness of rules, such as decision tree size or the compactness of classification rules.

**B. Association Rules:** The core purpose of the association rule analysis is to discover the interrelated and interdependent relationships in the data. Association rule mining is also a data mining method derived from database theory. Specifically, in relational databases, some data often appear synchronously, which is called a pattern. When this pattern appears frequently in the database, it is considered that there is a specific association relationship, which is called an association rule. Therefore, the revised content is determined to be consistent with the author's original intention [7].

The current research generally uses support and confidence as its measurement criteria. In the research process, scholars further strengthened the relevant parameters of association rules from different perspectives. It

incorporates the interest level and other indicators into the consideration range; thus, many new methods and applications have been proposed, and recent developments in association rule mining have also appeared. When specific patterns in the dataset meet the support and confidence thresholds, they become the association rules. First, the mining results need to reflect the closeness of the connections between the data. Data with fewer connections should not appear due to the association rules. Therefore, a reasonable minimum support threshold needs to be set. At the same time, we also pay attention to the credibility of the association rules and make appropriate requirements for credibility, which is reflected in the setting of the minimum confidence threshold.

**C. Clustering:** Compared with association rules, cluster analysis is another standard data mining method. It refers to a specific similarity estimation of the data to be analyzed in an unsupervised environment, and the data with higher similarity are combined, called clustering. The clustered result set has the characteristics of similarity of the same class and differences between classes, which is very suitable for grasping the distribution of data and its association. Clustering is the process of grouping a set of elements so that the elements in the same group or cluster are more similar than those in other groups or clusters. It is a common technique for statistical data analysis used in pattern recognition, information retrieval, bioinformatics, machine learning and image analysis. Clustering can be achieved by various algorithms that differ in the similarities required between elements of a cluster and how to efficiently find the elements of the clusters [43].

Most algorithms used for clustering try to create clusters with small distances among the cluster elements, intervals, dense areas of the data space or particular statistical distributions. In clustering, classes are unknown apriori and are discovered from the data.

Clustering algorithms in data mining are unsupervised Machine Learning algorithms comprising a set of data points in clusters so that the objects are precisely the same group. Clustering algorithms in data mining will help to split data into several subsets. Each subset has data like one another. Those subsets are called clusters; given that our client database is divided into clusters, we can decide who is most suitable for that item. Clustering algorithms in Data Mining help in the identification of aspects. That is similar to land use on an earth observation site. It can additionally aid in title groups of houses in a city. It is based on home geography, worth, and style location. Clustering algorithms in Data Mining likewise help classify documents online for data discovery. Also, we use Data clustering within outlier detection apps—for example, card fraud detection. As a data mining feature, cluster analysis serves as a tool. That is gaining insight into the distribution of data.

**D. Sequence:** Sequence pattern analysis is similar to association analysis, and its purpose is to dig out the connections between data, but the focus of sequence pattern analysis is to analyze the causal relationship between the data before and after.

## 2 Data mining technology

Data mining technology [45] has the following characteristics:

1. The amount of data is often tremendous. The so-called data mining must be built based on massive data, as small-scale data cannot reflect statistical laws. It isn't very meaningful to mine small-scale data, and the knowledge found is insufficient to reflect the actual situation in real life. It can be said that it is essentially wrong. In the face of massive amounts of data, it is imperative to reduce the time complexity of the related algorithms so that helpful knowledge and information can be mined effectively.
2. Potentially useful: The result of data mining should be an undiscovered rule or pattern which can provide detailed guidance for life and production. The results of excavations like "Many people holding umbrellas on rainy days" are meaningless.
3. Independence and indivisibility: The data mining process is highly complex and cannot be completed in a few simple steps. However, these steps cooperate and cannot complete the corresponding work independently. In this process, on the one hand, it is necessary to select specific algorithms to achieve mining efficiency; on the other hand, the relevant operators must have solid business skills. It can analyze and process data according to actual business needs, reasonably interpret the mining results, and correctly apply them to future work.

## 3 The steps of data mining

The four steps of data mining can be summarized as follows: (1) Data selection, (2) Data transformation, (3) Mining data, and (4) Interpreting the results.

Completing the data selection process obtains the specific partial data needed for data mining. Then, it is necessary to format the selected data further to provide an identifiable data input source for the following data mining step. After completing the two tasks of data selection and data format conversion, the next step is to use various data mining algorithms and integrated tools for the mining process. In the mining process, data warehouse and data mining algorithms are often used in combination. On the one hand, it reduces the calculation of specific statistical values; on the other hand, it reduces the consumption of extra time and space resources generated by data exchange. Although this method is commonly used, it does not restrict the algorithm from using the original data under appropriate conditions. In most cases, this approach is essential. The main advantage of using a data warehouse is that most of the data has been integrated into a suitable format, making it easier for data mining tools to extract high-quality information.

A reasonable interpretation of the mining results is the final step in the process. Through the steps, the mining results after analysis and processing are obtained. These mining results must be sent to the final decision-maker through the DSS in this step. Interpreting the mining results requires a reasonable interpretation and a deeper filtering of the data before sending it to the decision-making system. Once the mining results are unexplainable or unsatisfactory, the entire process must be repeated until beneficial results are produced.

In summary, data mining is not only an independent and indivisible process, but the realization of the process is also highly complicated. Many steps must be performed correctly before data can be provided to data mining tools. In addition, we cannot guarantee that the existing data mining tools will not produce meaningless results during the work process. Here, to a large extent, data mining is not a direct analysis operation on the original data but is based on a data warehouse, and the data warehouse provides a direct source of input data to the data mining tool. At the same time, the DSS tool will make the next step of processing the mining results. In this way, data mining will be combined with DSS to provide a final solution for enterprises to implement data mining strategies.

In general, the developers of data mining tools are also the people who pre-process the data. Therefore, a well-designed data mining tool will integrate related tools for data integration and format conversion. It is worth noting that although data mining uses data warehouses to provide processed data as input, it is not necessary in most cases. Instead, data can be downloaded directly from the operation file to an available file containing data that can be used for data mining and analysis.

Data mining technologies will deepen the economic statistics accumulated over a long period into the conditions data users require. Therefore, many characteristics of the data mining technology will be involved in the process of practice. According to these characteristics, ensure that economic statistics can play a role to the greatest extent and serve the needs of managers.

## 4 Data mining concepts

Data mining is a set of techniques for effectively and automatically detecting previously unknown, true, new, valuable and understandable trends in large databases. The results must be actionable for the organization's decision-making process. Traditionally used by business intelligence companies and financial analysts, it is primarily used in analysis to extract information from large data sets provided by new experimental and analytical methods [18]. Data mining strategies can be grouped as follows:

**A.Classification:** In this scenario, the data instance must be categorized into one of the identified or described target groups. Another consideration may be whether a consumer has to be listed as a trustworthy client or defaulter in a credit card transaction report, despite the different demographic and prior payment characteristics.

**B. Estimation:** Like arrangement, the motivation behind an estimation model is to decide an incentive for an obscure yield trait. Notwithstanding, in contrast to order, the yield characteristic for an estimation issue is numeric instead of straight out. For example, a model can be "Gauge the compensation of a person who possesses a games vehicle?"

**C. Prediction:** It is challenging to separate forecast from characterization or estimation. The main distinction is that the prescient model predicts a future result instead of determining the present conduct. The yield trait can be absolute or numeric. For example, a model can be "Anticipate one week from now's end cost for the Dow Jones Industrial Average", classifying a decision tree's development and its prescient applications.

**D. Association Rule Mining:** There, fascinating secret laws called association rules are mined in a massive transactional database. For example, the milk, butter→biscuit} rule provides information that if milk and butter are

bought together, biscuits are purchased in such a manner that they can be priced together to increase the total profits of each item.

**E. Clustering:** Clustering is a particular type of grouping in which the goal groups are not defined. E.g., 100 consumers have to be categorized based on specific similarity parameters, and it is not preconceived that the groups will eventually be divided.

## 5 Data mining models

Many standard models can be used successfully in several data mining issues. There are few decision trees, neural networks, Naive Bayes classifiers, Lazy trainers, Vector help devices, and regression-dependent classifiers. The most suitable model can be used depending on the type of use, the complexity of the data and the characteristics. There is still no straightforward answer to the question of which is the best data mining platform. One can only assume that one model is better for a particular application.

**A. Decision trees:** A decision tree is a flow-chart-like tree structure. Leaf nodes represent class labels or class distribution. A decision tree is a classifier in which each non-terminal node represents either a test or decision for the given data item. Which branch to select next depends on the test outcome [15]. To classify a given data item, start at the root node and follow the assertions until we reach a terminal node or leaf node. The decision tree is a common classification tool. It is a tree-like structure where each internal node reflects a judgement of the importance of the attribute. -the branch represents the result of the decision, and the tree leaves represent categories. The decision tree is a construction that is both analytical and concise. The decision tree shows the relationship in the training data. The decision tree method is a powerful statistical tool for classification, prediction, interpretation, and data manipulation that has several potential applications in medical research. An attribute in a learning problem may be nominal (categorical) or continuous (numerical). Numerical characteristics with a vast, even infinite domain become a fundamental challenge in pattern recognition, machine learning, and data mining. Mining data with numerical attributes require discretization before or throughout the model-building process. A special kind of discretization is performed through the decision tree construction process. The decision tree algorithm uses binarization which splits the numerical values into two intervals.

Using decision tree models to describe research findings has the following advantages:

- Simplifies complex relationships between input and target variables by dividing original input variables into significant subgroups.
- Easy to understand and interpret.
- Non-parametric approach without distributional assumptions.
- Easy to handle missing values without needing to resort to imputation.
- Easy to handle heavily skewed data without resorting to data transformation.
- Robust to outliers.

**B. Neural networks:** Neural networks [42] deliver a mathematical model which aims to imitate the human brain. Knowledge is referred to as a complex system of intertwined receptors called neurons. Every hub has a weighted association with different hubs in neighbouring layers. Singular hubs take the info from associated hubs and utilize the loads with a straightforward capacity to register yield esteems. Training in neural networks is developed by integrating weight changes, while some learning instances have been through the method more than once. When prepared, an obscure occasion going through the system is grouped by the qualities seen at the yield layer. Studies existing work on neural system development endeavouring to distinguish the significant issues included, bearings the work has taken and the present best in class.

Neural networks (NNs) more accurately called Artificial Neural Networks (ANNs). It is expressed in terms of the biological neuron system. It consists of many separate units. The individual units are communicated to each other by sending signals. It is similar to the brain, composed of many processing components. It is organized as a directed graph which contains nodes and the edges connecting each node. The advantages are the interconnections between each node. Consider the firing rate of every neuron. As said before, the neurons are interconnected. It receives 'm' inputs from 'n' nodes. Each edge connecting the node contains weight. The sum of the weights is calculated. The threshold value is assigned to each neuron. If the weighted sum exceeds the threshold value, it produces the output 1; otherwise, 0.

The topologies of Artificial Neural Networks are FeedForward and Feedback. In the FeedForward approach, the data flows only in one direction, so it does not receive any acknowledgement from the receiver side. Feedback cannot be sent in case of an indication of errors. This method is appropriate for recognizing images or identifying fingerprint patterns. The inputs and outputs are not changed. In the Feedback approach, it is possible to send the feedback. It is a little more efficient because of passing the indications then and there.

**C. Naive Bayes classifier:** The base for the naive Bayes classifier is the Bayes theorem. A hypothesis is generated for the given set of classes. In the naive Bayes algorithm, an independence assumption is made. Based on the target value, the attribute's values are chosen and independent [36]. The approach used in the naive Bayes classifier is straightforward. With the help of a small amount of training data, it is possible to classify the given instances. For example, to predict the fruit as an "apple", based on the colour red and its round shape, it is classified as an apple which shows it as an independent model.

This method is also suitable for complex situations. This classifier offers a straightforward yet excellent directed classification method. The model accepts all info credits as the equivalent significance and autonomy of each other. The naive Bayes classifier depends on the traditional Bayes hypothesis exhibited in 1763, which takes a shot at the likelihood hypothesis. In a fundamental phrase, the naive classifier of Bayes agrees that the proximity (or non-compliance) of a particular class variable is incompatible with the proximity (or non-appearance) of some other item.

Although these presumptions are likely false, the Bayes Classifier functions very well. Contingent to the accurate initiative of a probability model, naive Bayes classifiers can be produced in a regulated pick-up setting. In a variety of practical implementations, the parameter estimation for the naive Bayes model utilizes the most severe likelihood method.

**D. Association rules:** They are genuinely not unreasonably unique concerning classification decisions aside from that they can foresee any trait, not simply the class, allowing them to anticipate blends of qualities. Additionally, affiliation rules are not planned to be utilized altogether, as classification rules seem to be. Distinctive affiliation decides to express various regularities that underlie the dataset, and they, by and large, anticipate multiple things. Since such a large number of different affiliation rules can be gotten from even a modest dataset, intrigue is confined to those that apply to a sensibly considerable number of cases and have a sensibly high precision on the examples to which they apply to. The inclusion of an affiliation rule is the quantity of occurrences for which it predicts accurately. This is regularly called its support. Its exactness is regularly called certainty. It is the quantity of cases that it predicts effectively, communicated as an extent of all occurrences to which it applies [8].

**E. Machine learning and statistics:** Data mining can be considered as a juncture of measurements and AI. In truth, one ought not to search for a separating line between AI and insights because there is a continuum. Some get from the aptitudes instructed in standard measurement courses, and others are more firmly connected with the sort of AI that has emerged out of computer science. Verifiably, the different sides have had somewhat various customs. For example, while insights are progressively concerned about testing theories, AI is increasingly worried about detailing the procedure of speculation as a pursuit of potential ideas. In any case, this is a gross distortion.

## 6 Decision trees

Decision trees are one of the most popular methods for classification in various data mining applications. The primary learning approach of a decision tree is a greedy algorithm, which uses the recursive top-down approach of a decision tree structure. A decision tree is a tree-shaped schematic diagram used to determine a course of action or show a statistical probability. Each branch of the decision tree represents a possible decision or occurrence. The tree structure shows how one choice leads to the next, and using branches indicates that each option is mutually exclusive. Decision tree methodology is a commonly used data mining method for establishing classification systems based on multiple covariates or developing prediction algorithms for a target variable. This method classifies a population into branch-like segments constructing an inverted tree with a root node, internal nodes, and leaf nodes. The non-parametric algorithm can efficiently deal with large, complicated datasets without imposing a complex parametric structure. When the sample size is large enough, study data can be divided into training and validation datasets. Using the training dataset to build a decision tree model and a validation dataset to decide on the appropriate tree size needed to achieve the optimal final model.

During the construction of the decision tree, some uncovered errors may occur. These errors would be rectified initially; otherwise, it leads to the wrong decision-making process. Such errors can be corrected and evaluated using

an approach called tree pruning. There are two methods of pruning employed. One is pre-pruning, and the other is post-pruning. In pre pruning method, pruning is done in the initial stage. It checks for anomalies; if so, it stops the tree's construction at that stage itself. By halting, the node becomes a leaf node. In the post-pruning method, the entire tree is built. Then from the root, it starts pruning by removing the sub-trees. An individual sub-tree is taken from the original tree, and if it finds any anomaly, the corresponding branches are removed and replacing it with a leaf.

Decision tree types are based on the target variable we have. It can be of two types:

1. Categorical Variable Decision Tree: A decision Tree with a categorical target variable is called a Categorical variable decision tree.
2. Continuous Variable Decision Tree: A decision Tree has a continuous target variable called Continuous Variable Decision Tree.

## 7 Common usages of decision tree

Typical usages of decision tree [37] models include the following:

**A. Variable selection:** The number of variables routinely monitored in clinical settings has increased dramatically with the introduction of electronic data storage. Many of these variables are of marginal relevance and, thus, should probably not be included in data mining exercises. However, like stepwise variable selection in regression analysis, decision tree methods can be used to select the most relevant input variables that should be used to form decision tree models, which can subsequently be used to formulate clinical hypotheses and inform subsequent research.

**B. Assessing the relative importance of variables:** Once relevant variables are identified, researchers may want to know which variables play significant roles. Generally, variable importance is computed based on the reduction of model accuracy (or in the purities of nodes in the tree) when the variable is removed. In most circumstances, the more records a variable affect, the greater the importance of the variable.

**C. Handling of missing values:** A common but incorrect method of handling missing data is to exclude cases with missing values; this is both inefficient and risks introducing bias in the analysis. Decision tree analysis can deal with missing data in two ways: it can either classify missing values as a separate category that can be analyzed with the other types or use a built decision tree model which sets the variable with lots of missing values as a target variable to make a prediction and replace these missing ones with the predicted value.

**D. Prediction:** This is one of the most vital usages of decision tree models. Using the tree model derived from historical data, predicting the result for future records is easy.

**E. Data manipulation:** Too many categories of one categorical variable or heavily skewed continuous data are expected in medical research. In these circumstances, decision tree models can help decide how to best collapse categorical variables into a more manageable number of categories or subdivide heavily skewed variables into ranges. Building the decision tree is a recursive process. First, all the training data records are placed in the decision tree's top node or the root node. Next, the highest gain ratio attribute is chosen as the first splitting attribute. Once the data set is split into different nodes based on the first attribute, each node is then split based on the attribute with the highest gain ratio when applied to the data in the node. Finally, each node is given a majority label based on the class label of the majority of the records in the node.

The splitting is done by recursive partitioning, starting with all the observations represented by the node at the top of the tree [14]. The algorithm splits this parent node into two or more child nodes so that the responses within each child region are as similar as possible. The splitting process is then repeated for each child node, and the recursion continues until a stopping criterion is satisfied and the tree is fully built. At each step, the split is determined by finding the best predictor variable and the best cutpoint that assign the observations in the parent node to the child nodes. This process is continued until one of the three stopping conditions is met:

1. all of the records in the node belong to the same class;
2. the node is empty;
3. none of the attributes provide any further information gain.

Once a stopping condition is met, the final node is considered a leaf and is given a class label. Once every record in the data set is placed into a leaf, the decision tree-building process is complete. Fig. 1 illustrates a simple decision tree model with a single binary target variable Y (0 or 1) and two continuous variables, x1 and x2, ranging from 0 to 1. The main components of a decision tree model are nodes and branches.

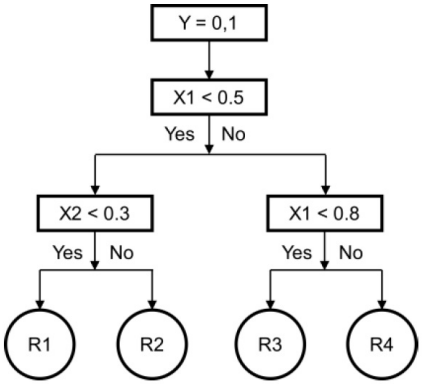


Figure 1: illustrates a simple decision tree.

## 8 Types of nodes

There are three types of nodes:

(a) A root node, also called a decision node, represents a choice that will result in subdividing all records into two or more mutually exclusive subsets.

(b) Internal nodes, also called chance nodes, represent one of the possible choices available at that point in the tree structure; the top edge of the node is connected to its parent node, and the bottom edge is connected to its child nodes or leaf nodes.

(c) Leaf nodes, also called end nodes, represent the final result of a combination of decisions or events

**A. Branches:** Branches represent chance outcomes or occurrences from the root and internal nodes. A decision tree model is formed using a hierarchy of branches. Each path from the root node through internal nodes to a leaf node represents a classification decision rule. These decision tree pathways can also be described as 'if-then' rules. For example, "if condition one and condition two and condition ... and condition k occur, then outcome j occurs."

**B. Splitting:** Only input variables related to the target variable are used to split parent nodes into purer child nodes of the target variable. Discrete and continuous input variables (collapsed into two or more categories) can be used. When building the model, one must first identify the most crucial input variables and then split records at the root node and subsequent internal nodes into two or more categories or 'bins' based on the status of these variables. Characteristics related to the degree of 'purity' of the resultant child nodes. the proportion with the target condition) These characteristics are used to choose between different potential input variables, including entropy, Gini index, classification error, information gain, gain ratio, and towing criteria. This splitting procedure continues until pre-determined homogeneity or stopping criteria are met. In most cases, not all potential input variables will be used to build the decision tree model, and in some cases, a specific input variable may be used multiple times at different levels of the decision tree.

**C. Stopping:** Complexity and robustness are competing characteristics of models that must be simultaneously considered when building a statistical model. The more complex a model is, the less reliable it will be when used to predict future records. An extreme situation is to build a very difficult decision tree model that spreads wide enough to make the records in each leaf node 100% pure (i.e., all records have the target outcome). Such a decision tree would be overly fitted to the existing observations and have few records in each leaf, so it could not reliably predict future cases and, thus, would have poor generalizability (i.e., lack robustness). To prevent this, stopping rules must be applied when building a decision tree to prevent the model from becoming overly complex. Standard parameters used in controlling rules include:

1. the minimum number of records in a leaf;
2. the minimum number of records in a node before splitting;
3. the depth (i.e., number of steps) of any leaf from the root node. Stopping parameters must be selected based on the analysis's goal and the dataset's characteristics.

**D. Pruning:** In some situations, stopping rules do not work well. An alternative way to build a decision tree model is to grow a large tree first and then prune it to optimal size by removing nodes that provide less additional information. A standard method of selecting the best possible sub-tree from several candidates is to consider the proportion of records with error prediction (i.e., the ratio in which the predicted occurrence of the target is incorrect). Other methods of selecting the best alternative are to use a validation dataset (i.e., dividing the sample in two and testing the model developed on the training dataset on the validation dataset) or, for small samples, cross-validation (i.e., dividing the sample into ten groups or 'folds', and testing the model developed from 9 folds on the 10th fold, repeated for all ten combinations, and averaging the rates or erroneous predictions). There are two types of pruning, pre-pruning (forward pruning) and post-pruning (backward pruning). Pre-pruning uses Chi-square tests [4] or multiple-comparison adjustment methods to prevent the generation of non-significant branches. Post-pruning is used after generating a full decision tree to remove branches in a manner that improves the accuracy of the overall classification when applied to the validation dataset.

In machine learning, classification is supervised learning with many practical applications, from catching spam emails to categorizing tumour scans. Classification problems analyze data sets containing a collection of records that each have a bunch of attributes and a class label. The task is to create a model that maps each record's feature set onto its class label. A classification model can be used for descriptive purposes by summarizing the attributes in a data set that correlate with a specific class label. It can also be used for predictive purposes by classifying new records with unknown class labels. There are many different ways to create a classification model based on a data set, but they all follow a similar approach. First, the data set must be split into a set of training data and a set of testing data. The training data is composed of records where the class label is included. Each classification technique applies a different learning algorithm to the training data to build the classification model.

Once the model is constructed, it is then applied to the testing data, composed of records where the class label is removed. The model's accuracy can then be calculated by comparing the class labels predicted by the model to the actual class labels of the testing data. There are two general types of attributes in a data set. Discrete details are composed of values from a finite or countably infinite set, such as the set of natural numbers or non-numeric values. Continuous attributes are composed of values from an uncountably infinite set, such as the set of real numbers, which includes all decimal values. Many learning algorithms can only use discrete attributes. Thus, these algorithms must employ different methods to discretize continuous attributes.

One of the well-known decision tree algorithms is C4.5 [1]; C4.5 was developed by Quinlan Ross, which is an extension to ID3 [19]. It is mainly used for generating a decision tree. The main goal of the algorithm is to discover relationships between the values of a target or dependent attribute and those of some independent features. The improvements of C4.5 include:

1. employ information gain ratio instead of information gain. The gain ratio measure is a selection criterion used less biased towards selecting attributes with more values.
2. it is suitable for handling both categorical and continuous data. A threshold value is fixed such that all the values above the threshold are not considered.
3. handling incomplete training data with missing values. There are several ways to take missing attributes. Some of these are Case Substitution, Mean Substitution, Hot Deck Imputation, Col Deck Imputation, and Nearest Neighbour Imputation [26]. In addition, C4.5 can be directly used to predict missing attribute values. This is done by using the target attribute values to be predicted for test cases, as the classes used for training. The training data should therefore have the target attribute value specified.
4. prune during the construction of trees to avoid overfitting. There are two different pruning approaches: preprunning and postpruning. In pre-pruning, a decision tree is halted while growing so it won't get too complex. However, in post-pruning, the tree is developed to its fullest and then pruned following a bottom-up or a top-down strategy. Overfitting can be avoided by a stopping criterion that prevents some sets of training cases from being subdivided or by removing some of the structure of the decision tree after it has been produced.

## 9 The C4.5 tree-construction algorithm

A classification tree is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions and then splitting it up further on each branch. Initially, all objects are considered as a single group. The group is divided into two subgroups using a criteria, say high values of a variable for one group and low values for the other. The two subgroups are then split using the values of a second variable. The splitting process continues until a suitable stopping point is reached. The values of the splitting variables can be ordered or unordered categories. A classification tree is one of the most common decision tree types, in which the objective is to find the tree that best discriminates between classes. The decision tree generally represents a set of splitting rules organized in levels in a flowchart structure [35]. The classification process has two phases: the first phase is the learning process, where the classification algorithm analyzes the training data. The learned model or classifier is represented in the form of classification rules. The second phase is the classification process, where the test data are used to estimate the accuracy of the classification model or classifier. Fig. 2. provides a classification process in data mining.

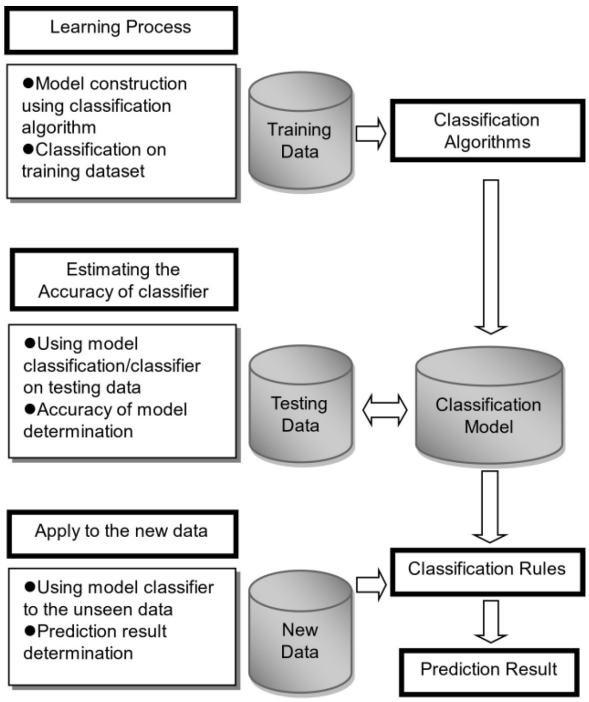


Figure 2: Classification Process in Data Mining

## 10 Decision trees structure

A decision tree is a data structure consisting of decision nodes and leaves. Decision trees solve classification problems in various fields such as engineering, science, medical, and other areas. Thus, decision trees are considered one of the most powerful tools to accomplish classification. Additionally, the tree concept is easily understandable compared to other classification methods and can also be used by other scientific field researchers than computer science. A decision tree structure is made of root, internal and leaf nodes. The tree structure is used in classifying unknown data records. At each internal node of the tree, a decision tree of the best split is made using impurity measures.

Two major phases should be ensured in decision trees: (1) Building the tree. (2) Classification. This process will be repeated until a leaf is encountered. The instance is then classified in the same class as the one characterizing the reached leaf. Several algorithms have been developed to ensure the construction of decision trees and their use for the classification task. The C4.5 algorithm is probably the most popular ones.

## 11 The C4.5 algorithm

The C4.5 algorithm is essentially an extension of the ID3 algorithm for decision tree learning. It can also handle continuous attributes and attributes with missing values [6]. The tree-growing process in the case of C4.5 is the same as that of ID3, finding the best split at each node using the information gain measure. But in the case of the continuous attribute, the C4.5 algorithm has to perform the additional step of converting it to a two–value categorical attribute by splitting about a suitable threshold. This threshold is chosen so that the resultant split produces maximum information gain. In C4.5, each node in a tree is associated with a set of cases. The decision tree has root nodes, internal nodes and leaf nodes.

At the beginning, only the root is present. This can be considered as the starting point of the tree where there are no incoming edges but zero or more outgoing edges. An internal node appears after a root node or an internal node and is followed by either internal or leaf nodes. It has only one incoming edge and at least two outgoing edges. Internal nodes are always attributes of the decision tree model. Leaf nodes are the bottom element of the tree and typically represent the class of the decision tree model. Depending on the situation, a leaf node might not always mean a class label because sometimes a decision cannot be made for some leaves. In that case, those leaves can be marked with signs such as a question mark. However, if it can be classified, each leaf node can have only one class label or sometimes a class distribution. Leaf nodes have one incoming edge and no outgoing edges. The standard C4.5 algorithm splits continuous attributes into two groups by choosing one split point with the minimum expected information requirement.

However, this process is often inaccurate and inefficient. The two groups will likely provide little information gain because each value is not likely to be closely related. Attributes with low information gain are detrimental to the model's overall accuracy. Furthermore, iterating through every possible split point in a large data set and calculating the expected information required for each point can consume large amounts of time.

In a nutshell, the C4.5 algorithm is implemented recursively with the following sequence:

1. Choose an attribute for the root node.
2. Create a branch for each value of that attribute.
3. Split cases according to branches.
4. Repeat the process for each branch until all cases have the same class.
5. Choosing which attribute to be used as a root is based on the highest gain of each attribute.
6. Return tree.

## 12 C4.5 algorithm measures

**A. Entropy:** In information theory, entropy measures a data file's uncertainty among random variables. Claude E. Shannon has developed the idea of entropy of random variables. He introduced the beginnings of information theory and the modern Ergodic theory age. Entropy and related information provide the long-term behaviour of random processes useful for data analysis. The behaviour of random processes is also a key factor for developing the coding for information theory. Entropy is a measure of average data collection uncertainty when we do not know the outcome of an information source. That means it measures how much information we do not have. This also indicates the average amount of information we will receive from the outcome of an information source [2]. Entropy is defined in equation (12.1).

$$\text{Info}(s) = \sum_{i=1}^{n} P_i \times \log_2^p \tag{12.1}$$

where $p_i$ is the probability that any subset of S belonging to category $C_i$, and $p_i = |C_i|/|S_i|$.

**B. Information Gain:** C4.5 uses information gain as its attribute selection measure. The attribute with the highest information gain is chosen as the splitting attribute. This attribute minimizes the information needed to classify the tuples in the resulting partitions and reflects the least randomness or impurity in the partitions. The information gain, G (S,A) of an attribute A, relative to the collection of examples S, is defined in equation (12.2):

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^{n} \frac{(|sv|)}{(|s|)}, \tag{12.2}$$

where value $(A)$ is the set of all possible values for attribute $A$, and $SV$ is the subset of $S$ for attribute $A$. Research shows equation (12.2) tends to choose the attribute with many values as the branch node. However, the attribute with many values is not always the best.

**C. Gain Ratio:** The information gain is good when it is used for a small or medium number of values. The notion of information gain introduced earlier tends to favour attributes that have a large number of values. C4.5 uses an extension of information gain known as the gain ratio, which attempts to overcome this bias. The split info can be defined in equation (12.3):

$$\text{Split Info}(S, A) = \sum_{i=1}^{v} \frac{|S_t|}{|S|} \times \log_2 \frac{|S_t|}{|S|}. \tag{12.3}$$

The first term in the equation for gain is just the entropy of the original collection S and the second term is the expected value of the entropy after Sis partitioned by this second term is simply the sum of the entropies of each subset weight by the fraction of examples $|S_t|/|S|$, that belong to Gain (S,A) is, therefore, the expected reduction in entropy caused by knowing the value of attribute A.

Split information gives positive results to most applications, but it has substantial limitations when used in decision trees. Although information gain is usually a good measure for deciding the relevance of an attribute, the result is not guaranteed for all attributes. One limitation of gain is that it tends to favour features with a large number of values that split the data into many small subsets. Quinlan suggested using a Gain Ratio instead of Gain to compensate for this. The Gain Ratio is defined in equation (12.4).

$$\text{Gain Ratio}(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)}. \tag{12.4}$$

The attribute with the maximum gain ratio is selected as the splitting attribute. However, the gain ratio becomes unstable when the split information is equal to zero or close to zero [36].

Based on the information gain and gain ratio study, the remarks of information entropy in C4.5 is discussed below.

1. It gives inferior results when large distinct values are used in discrete and continuous attributes.
2. No specific measurement technique is available to predict actual information gain before it is applied. Therefore, the information gain ratio is evaluated only after the generation of attribute values. As a result, the mismatch or wrong selection of attributes may give less performance and accuracy.
3. It fails When the information gain is less than the number of attributes used.
4. One of the critical issues in decision tree information entropy is uncertainty. If the previously chosen attribute has less value, then it is more complex to select the next. It leads to an unconditional selection of attributes.

## 13 Rule sets

Generally, rules are perfect ways to express knowledge or represent information acquired from a plain or mined data set. For example, in association rule mining, inferences are made from data and defined as rule sets. Nonetheless, another use for rules exists in decision trees. Decision trees are preferred because of their simplicity in interpreting the results of the processed dataset [10]. However, in most real-world cases, decision trees can grow into very big and complex structures, making them hard to interpret even though they might have been pruned. Hence, rules come in handy when the trees are simplified into IF-THEN rules, referred to as decision rules.

An IF-THEN rule is an expression in the form; an IF condition THEN conclusion. The "IF" part or the left-hand side of a rule is called the rule antecedent or precondition. The "THEN" part or the right-hand side of a rule is called the rule consequent. The consequence of the rule withholds the classification label or the prediction of that rule. Extracting rules from a grown decision tree is very straightforward. Each extracted rule is the path from the root to a leaf node. To extract rules, each split attribute is "AND" to the "IF" part of the rule according to its value until the leaf node is reached, which forms the consequent or the "THEN" part of the rule. The rules extracted are mutually exclusive and exhaustive since they are directly extracted from the tree.

A disjunction or "OR" implication between the extracted rules supports the idea of mutual exclusiveness. This also means the rules cannot overlap or conflict with each other since the extracted rules match the leaves of the tree in a one-on-one relationship. The exhaustive term implies a rule for each training case (for each attribute-value

combination occurring in the tree). A rule set or decision rules can also be pruned like a decision tree. Sometimes the rules or parts can be useless or have poor decision tree accuracy since the rules might have been extracted from an unpruned tree. In this case, the rules are pruned according to some pruning algorithm. For example, the C4.5 algorithm has a feature where it produces decision rules as well as a decision tree and prunes the rules using error-based pruning. A consequence of rule pruning might be losing mutually exclusive property since, after pruning, there will not be any guarantee that each possible path will go to a separate leaf node.

## 14 Related works

Researchers have offered several techniques to improve the C4.5 algorithm. In this section, we are going to discuss a few recent works.

Gaurav and Hitesh proposed the C4.5 algorithm, which is enhanced by using the L 'Hospital Rule; this simplifies the calculation process and improves the efficiency of decision-making algorithms. Ploat and Gunes have offered a 'one against all approach' with C4.5. They have experimented with three famous data sets: Dermatology and Image segmentation from UCI [5]. In their experiment, they found excellent accuracy against other algorithms.

Jiang and Yu [9] have proposed a hybrid algorithm based on outlier detection and C4.5. They have worked with imbalanced data to make them balance using outlier detection; they implement the C4.5 algorithm. Their proposed algorithm shows good accuracy relative to other algorithms, namely C4.5 and Ripper.

Yu and Ai [44] have worked on classifying Remote Sensing (RS) data using a rough set and C4.5 algorithm. Their algorithm performs well on that specific data type.

Yang [41] has used hierarchical clustering to limit the decision tree to a binary tree to improve the traditional C4.5 algorithm. The author's algorithm successfully timed down the number of leaf nodes and enhanced accuracy.

Researchers [3] from India used the C4.5 algorithm on a web-based Soya Bean Expert System. The proposed Bagging algorithm was used to improve the performance of C4.5. As a result, researchers could enhance the performance of C4.5 between 4% to 6%.

Carlos J. Mantas and Joaquin Abellan [13] have presented a new model called Credal-C4.5, a modified version of the C4.5 algorithm. It has been defined by using a mathematical theory of imprecise probabilities and uncertainty measures on creedal sets. They have shown C4.5, and Credal-C4.5 are very similar in performance when no noise is added, and the only difference is that Credal-C4.5 has a better performance than C4.5 and, in this case, also the number of nodes of Credal-C4.5 is notably lower.

Some researchers [34] studied the classification of blood characteristics by a C4.5 decision tree, a naïve Bayes classifier and a multilayer perceptron for thalassaemia screening. They aimed to classify eighteen classes of thalassaemia abnormality, which have a high prevalence in Thailand and one control class by inspecting data characterized by a complete blood count (CBC). Their experiment involving stratified 10-fold cross-validation revealed that the naïve Bayes classifier and multilayer perceptron are the most suitable classifier for the data pre-processed by attribute discretization.

## 15 The proposed measure

One of the features of information gain is the tendency to choose the attribute with the most values, which is considered a defect. Therefore, to overcome this defect, Quinlan has represented the gain ratio formula. The gain ratio formula places split info in the fraction's denominator. However, if its value becomes close to zero or equal to zero, it causes instability of the gain ratio and therefore decreases the accuracy of the decision tree. Instead of using split info in gain ratio formula's denominator, this article proposes using split information and the number of values of the attribute (N) summation. So that not only the instability of the gain ratio is eliminated but also the accuracy of the decision tree is significantly increased. The improved gain (S,A) of an attribute A, relative to a set of training instances S, can be defined by equation (15.1)

$$\text{Improved Gain}(S, A) = \frac{\text{Gain}(S, A)}{\text{Split Info}(S, A) + N} \tag{15.1}$$

where Gain(S,A) is the information gain that the attribute A partitions the training instances S. Split information is the potential information generated by splitting the training dataset S and N is the number of values of the attribute A.

This experiment's hardware environment is CPU Core5 P9600, Memory 6G, Hard disk 700G, and Windows 7 system. We have evaluated the formula with experiments on 4 data sets. All these data sets are picked up from the popular UCI (University of California at Irvine) data repository. Table 1 shows detailed information on those data files.

Table 1: Data sets properties

| **Data set Name** | **Specifications Data Sets** | | | |
|---|---|---|---|---|
| | **Number of Instances** | **Number of Attributes** | **Attributes Type** | **Miss Value** |
| Mammographic Mass | 820 | 6 | Discrete | no |
| Breast Cancer | 681 | 10 | and | no |
| Liver | 345 | 7 | Continuous | no |
| Tea | 125 | 6 | discrete | no |

## 16 Classifiers accuracy

When a model is built for a decision tree, the first question that comes into mind is how accurate or reliable the model is on unseen cases. This is why evaluating decision trees is essential because one should be sure that the resulting decision tree will be reliable and efficient. In some cases, there might be more than one decision tree model for a specific machine learning problem, and one must be preferred over the others. In such cases, the only option to overcome such a problem is to take some precautionary steps. This is achieved by using measures and metrics to estimate the overall performance of the inducer's model for future use. Before discussing what kind of measures there are, the metrics used for performance evaluation need to be explained. A metric for decision tree performance can have various meanings. In some cases, the performance is measured by speed, sometimes by the size of the grown tree and in most cases, it is measured by accuracy. Below are some metrics that have been considered viable and their definitions:

## 17 Accuracy based

These measures show the performance of classifiers on rating systems or percentages. Accuracy [38] based metrics have dominated the evaluation methods and techniques since they give the most realistic and easily calculable results. Some are accuracy (recognition rate), error rate, recall, specificity and precision.

**A. Speed:** It is usually referred to as the computational costs that are encountered during building the model and using it afterwards.

**B. Robustness:** This is how reliable or correct predictions a classifier makes when it encounters noisy data or data with missing values.

**C. Scalability:** This can be considered an aspect to evaluate when the classifier is given large amounts of data. It measures how well the classifier operates given a large amount of data and is usually assessed by classifying data of increasing size.

**D. Interpretability:** The amount or extent to which the classifier's results can be interpreted. This measurement can be tough to assess different classifiers based on it since it is subjective. As mentioned earlier, the interpretability of decision trees can be easy until some point; however, it is 38 inevitable that it might become tough to interpret if the tree becomes complex. Now that the important metrics have been identified, measures in the accuracy-based actions can be explained.

Previously, metrics and measurements required in accuracy estimating methodologies were discussed to build a clear connection between the two concepts. As mentioned earlier, accuracy is the most suitable measure for the performance evaluation of decision trees. Consequently, all the estimating methodologies discussed are based on accuracy metrics, hit rates, error rates, etc. Estimating accuracy is vital for several reasons. Firstly, verifying if a model is reliable for future predictions is necessary. Secondly, when there is more than one model, there needs to be some measurement or a metric that can separate the best among multiple models, and this is where an accuracy estimation method comes

in. Lastly, it can be used to assign confidence intervals to multiple inducers to optimize the outcome of a combining inducer.

The accuracy measure represents how far the tuples are being classified correctly. Usually, the algorithm's performance is examined by evaluating the accuracy of the result. The accuracy of a decision tree is computed using a testing data set. Two-fold cross-validation is applied in this work to assess the performance of a decision tree. The basic process is: a complete data set is split into two parts, one part of the dataset being dedicated to the training and the other for the testing. The training set is used to learn the algorithm and generate the tree, while the testing set is used to evaluate the generated decision tree.

Table 2: The model accuracy for data sets.

| **Table Head** | **Accuracy (%)** | |
|---|---|---|
| | **C4.5** | **IC4.5** |
| Mammographic Mass (MM) | 70% | 80% |
| Breast Cancer | 72% | 85% |
| Liver | 67% | 74% |
| Tea | 73% | 82% |

This procedure is repeated; hence, every part of the data set is used for testing and training data. Afterwards, the overall accuracy parameters are calculated as means for the evaluation of the individual cross-validation subset. A high accuracy signifies that the testing process is working well with a valid theory. Table II shows the model accuracy in percent for two classifications (C4.5 and IC4.5) and Fig. 3 shows a comparison of the classifiers' accuracy. It shows clearly that IC4.5 Algorithm has more precision than C4.5 Algorithm.

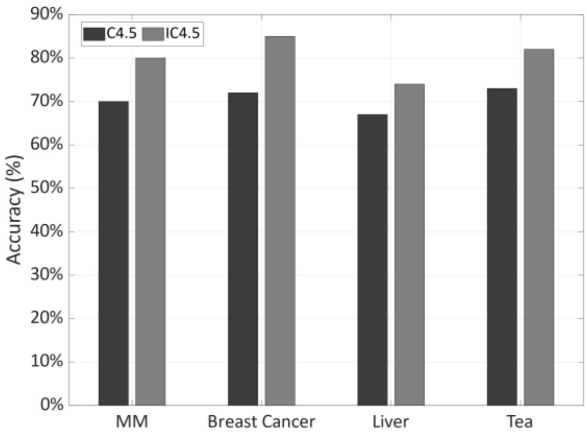


Figure 3: Comparison of classifiers Accuracy

## 18 Conclusion

The C4.5 algorithm makes it possible to classify the data sets that have quantitative attributes. In addition, it is possible to handle missing values with this algorithm and both continuous and discrete attributes. C4.5 creates a threshold to conduct continuous attributes and splits the list into those whose attribute value is above the threshold and those less than or equal. The C4.5 algorithm performs well in constructing decision trees and extracting rules from the data set. This work aims to increase the classification accuracy to build a classification model. We proposed the IC4.5 algorithm instead of the C4.5 decision tree. Different kinds of data sets are used as input to calculate the accuracy of the IC4.5 and C4.5 algorithms. The results obtained from experiments show that IC4.5 is better than C4.5 in increasing the accuracy of the decision tree.

## References

[1] L.I.P. Aji and A. Sunyoto, *An implementation of C4. 5 classification algorithm to analyze student's performance*, 3rd Int. Conf. Inf. Commun. Technol. (ICOIACT), IEEE, 2020, pp. 126–130.

[2] B.L. Aurelian, *An information entropy based splitting criterion better for the data mining decision tree algorithms*, 22nd Int. Conf. Syst. Theor Control Comput. (ICSTCC), IEEE, 2018, pp. 535–540.

[3] M.S.P. Babu, S.R.S. Reddy, B.V. Ramana, and N.V.R. Murty, *A web-based soya bean expert system using bagging algorithm with C4.5 decision trees*, Int. J. Agricul. Innov. Res. **1** (2020), no. 4, 2319–1473.

[4] C.-C. Chang, C.-H. Huang, and Y.-S. Chu, *A hardware-friendly pruning approach by exploiting local statistical pruning and fine grain pruning techniques*, IEEE Int. Conf. Consumer Electr.-Asia (ICCE-Asia), IEEE, 2022, pp. 1–3.

[5] S. Chang, Y. Shihong, and L. Qi, *Clustering characteristics of UCI dataset*, 39th Chinese Control Conf. (CCC), IEEE, 2020, pp. 6301–6306.

[6] P. Chen, *The application of an improved C4. 5 decision tree*, 7th Ann. Int. Conf. Network Inf. Syst. Comput. (ICNISC), IEEE, 2021, pp. 392–396.

[7] S. Datta and K. Mali, *Significant association rule mining with high associability*, 5th Int. Conf. Intell. Comput. Control Syst. (ICICCS), IEEE, 2021, pp. 1159–1164.

[8] G. Ding, *Optimization model of physical education teaching evaluation based on association rule algorithm*, 12th Int. Conf. Intell. Comput. Technol. Autom. (ICICTA), IEEE, 2019, pp. 516–519.

[9] S. Jiang and W. Yu, *A Combination Classification Algorithm Based on Outlier Detection and C4.5*, Springer Publications, 2019.

[10] S. Kottam and V. Paul, *Multi-soft set based approach for rule based classification using sequential covering algorithm*, Adv. Comput. Commun. Technol. High Perform. Appl. (ACCTHPA), IEEE, 2020, pp. 34–42.

[11] D.T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, John Wiley & Sons, 2005.

[12] O.Z. Maimon and L. Rokach, *Data Mining with Decision Trees: Theory and Applications*, **81**, World Scientific, 2014.

[13] C.J. Mantas and J. Abellan, *Credal-C4. 5: Decision tree based on imprecise probabilities to classify noisy data*, Expert Syst. Appl. **41** (2014), no. 10, 4625–4637.

[14] D.P. Mishra, I. Goyal, and S.K. Behera, *A comparative study of two different octagonal structure-based split ring resonators*, IEEE Microwaves Antennas Propag. Conf. (MAPCON), IEEE, 2022, pp. 1007–1012.

[15] S. Mitrofanov and E. Semenkin, *An approach to training decision trees with the relearning of nodes*, Int. Conf. Inf. Technol. (InfoTech), IEEE, 2021, pp. 1–5.

[16] B.M. Mosammam, N. Rasekh, M. Mirsalim, and A. Khorsandi, *Electromagnetic analysis for dd pad magnetic structure of a wireless power transfer (WPT) for electrical vehicles*, Smart Grid Conf. (SGC), IEEE, 2018, pp. 1–6.

[17] B.M. Mosammam, N. Rasekh, M. Mirsalim, and J. Shokrollahi Moghani, *Comparative analysis of the conventional magnetic structure pads for the wireless power transfer applications*, 10th Int. Power Electr. Drive Syst. Technol. Conf. (PEDSTC), IEEE, 2019, pp. 624–628.

[18] J. Peng, J. Guo, Q. Yang, J. Lu, and J. Shao, *A general framework for mining concept-drifting data streams with evolvable features*, IEEE Int. Conf. Data Min. (ICDM), IEEE, 2021, pp. 1276–1281.

[19] A. Rajeshkanna and K. Arunesh, *ID3 decision tree classification: An algorithmic perspective based on error rate*, Int. Conf. Electron. Sustain. Commun. Syst. (ICESC), IEEE, 2020, pp. 787–790.

[20] N. Rasekh, S. Dabiri, N. Rasekh, M. Mirsalim, and M. Bahiraei, *Thermal analysis and electromagnetic characteristics of three single-sided flux pads for wireless power transfer*, J. Clean. Prod. **243** (2020), 118561.

[21] N. Rasekh, J. Kavianpour, and M. Mirsalim, *A novel integration method for a bipolar receiver pad using LCC compensation topology for wireless power transfer*, IEEE Trans. Vehic. Technol. **67** (2018), no. 8, 7419–7428.

[22] N. Rasekh and M. Mirsalim, *Evaluation study on an integration method for a DDQP using LCC and series compensation topologies for inductive power transfer*, IET Electric Power Appl. **12** (2018), no. 9, 1320–1327.

[23] N. Rasekh and M. Mirsalim, *Analysis of a compact and efficient DDQ pad integrated to the LCC compensation*

*topology for IPT*, 9th Ann. Power Electr. Drives Syst. Technol. Conf. (PEDSTC), IEEE, 2018, pp. 26–29.

[24] N. Rasekh and M. Mirsalim, *Design of a compact and efficient Bipolar pad with a new integration of LCC compensation method for WPT*, 9th Ann. Power Electr. Drives Syst. Technol. Conf. (PEDSTC), IEEE, 2018, pp. 44–47.

[25] N. Rasekh, N.Rasekh, and M. Mirsalim, *Evaluation study of different integration methods of LCC compensation network for various types of magnetic structures of wireless power transfer*, 29th Iran. Conf. Electr. Engin. (ICEE), IEEE, 2021, pp. 284–289.

[26] N. Rasekh, J. Wang, and X. Yuan, *A new method for offline compensation of phase discrepancy in measuring the core loss with rectangular voltage*, IEEE Open J. Ind. Electr. Soc. **2** (2021), 302–314.

[27] N. Rasekh, J. Wang, and X. Yuan, *A novel in-situ measurement method of high-frequency winding loss in cored inductors with immunity against phase discrepancy error*, IEEE Open J. Ind. Electr. Soc. **2** (2021), 545–555.

[28] N. Rasekh, J. Wang, and X. Yuan, *In-situ measurement and investigation of winding loss in high-frequency cored transformers under large-signal condition*, IEEE Open J. Ind. Appl. **3** (2022), 164–177.

[29] N. Rasekh, J. Wang, and X. Yuan, *Artificial neural network aided loss maps for inductors and transformers*, IEEE Open J. Power Electron. **3** (2022), 886–898.

[30] N. Rasekh, J. Wang, and X. Yuan, *Variation of core loss in high-frequency transformers under different load conditions*, 24th Eur. Conf. Power Electron. Appl. (EPE'22 ECCE Europe), Hanover, Germany, **24** (2022), pp. 1–10.

[31] N. Rasekh, J. Wang, and X. Yuan, *Towards standardized magnetic components under power electronics excitation with user-friendly loss maps*, IEEE J. Emerg. Selected Topics Power Electronics, 2023.

[32] N. Rasekh, J. Wang, and X. Yuan, *Variation of core loss in high-frequency transformers under different load conditions*, IEEE Open J. Ind. Electr. Soc. **4** (2023), 316-327.

[33] S. Rezazade, A. Shahirinia, R. Naghash, N. Rasekh, and S.E. Afjei, *A novel efficient hybrid compensation topology for wireless power transfer*, IEEE Trans. Ind. Electron. **70** (2022), no. 3, 2277–2285.

[34] D. Setsirichok, T. Piroonratana, W. Wongseree, T. Usavanarong, N. Paulkhaolarn, C. Kanjanakorn, M. Sirikong, C. Limwongse, and N. Chaiyaratana, *Classification of complete blood count and haemoglobin typing data by a C4. 5 decision tree, a naïve Bayes classifier and a multilayer perceptron for thalassaemia screening*, Biomed. Signal Process. Control **7** (2012), no. 2, 202–212.

[35] A. Sharma, M.K. Sharma, and R.K. Dwivedi, *Improved decision tree classification (IDT) algorithm for social media data*, 10th Inte. Conf. Syst. Model. Adv. Res. Trends, IEEE, 2021, pp. 155–157.

[36] Y. Sugiarti, A. Supriyatna, I. Carolina, R. Amin, and A. Yani, *Model Naive Bayes classifiers for detection apple diseases*, 9th Int. Conf. Cyber IT Serv. Manag., IEEE, 2021, pp. 1–4.

[37] A.A. Supianto, A.J. Dwitama, and M. Hafis, *Decision tree usage for student graduation classification: A comparative case study in the faculty of computer science Brawijaya University*, Int. Conf. Sustain. Inf. Engin. Technol., IEEE, 2018, pp. 308–311.

[38] A.A. Tyugashev, D.N. Frantasov, and Y.V. Kudryashova, *Methods and tools of electricity metering information and measurement system accuracy increasing*, Int. Conf. Ind. Engin. Appl. Manufact., IEEE, 2019, pp. 1–5.

[39] J. Wang, N. Rasekh, X. Yuan, and K.J. Dagan, *An analytical method for fast calculation of inductor operating space for high-frequency core loss estimation in two-level and three-level PWM converters*, IEEE Trans. Ind. Appl. **57** (2020), no. 1, 650–663.

[40] J. Wang, X. Yuan, and N. Rasekh, *Triple pulse test (TPT) for characterizing power loss in magnetic components in analogous to double pulse test (DPT) for power electronics devices*, IECON 46th Ann. Conf. IEEE Ind. Electronics Soc., IEEE, 2020, pp. 4717–4724.

[41] C.C. Wu, Y.L. Chen, Y.H. Liu, and X.Y. Yang, *Decision tree induction with constrained number of leaf node*, Appl. Intell. **45** (2016), 673–685.

[42] R. Xin, J. Zhang, and Y. Shao, *Complex network classification with convolutional neural network*, Tsinghua Sci.

Technol. **25** (2020), no. 4, 447–457.

[43] H. Xu, *Research on clustering algorithms in data mining*, 3rd Int. Conf. Big Data Artific. Intell. Internet of Things Engin. (ICBAIE), IEEE, 2022, pp. 652–655.

[44] M. Yu and T.-H. Ai, *Study of RS data classification based on rough sets and C4.5 algorithm*, Int. Symp. Spat. Anal. Spat.-Temp. Data Model. Data Min., SPIE, 2009, pp. 95–102.

[45] K. Yu, T. Yuan, and Y. Li, *Application of data mining technology in sports data analysis in colleges and universities*, Int. Conf. Inf. Technol. Contemp. Sports, IEEE, 2021, pp. 329–332.