Comparison of Various Classification Algorithm on IRIS Dataset using WEKA

**Index**

|  |  |
| --- | --- |
| Abstract…………………………………………………………………. | 2 |
| Introduction………………………………………………………….. | 2 |
| Dataset…………………………………………………………………. | 2 |
| Result………………………………………………………………….... | 3 |
| Comparison between decession tree and SVM …… | 4 |
|  |  |
|  |  |
|  |  |

**Abstract:**

Classification is one of the most important task of Machine Learning. Main task of machine learning is data analysis. For study purpose various algorithm available for classification like decision tree, Navie Bayes, Back propagation, Neural Network, Artificial Neural, Multi-layer perception, Multi class classification, Support vector Machine, k-nearest neighbor etc. In this paper we introduce Three algorithms from them (Decision Tree, Support vector Machine, Apriori algorithm). Study purpose we take iris.arff dataset. Implement this all algorithm in iris dataset and compare between them. Weka is inbuilt tools for machine learning. So we used WEKA for implementation.

**Introduction:**

Generally machine learning is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both. Machine learning algorithms which carry out the assigning of objects into related classes are called classifiers. Classification algorithms include two main phases; in the first phase they try to find a model for the class attribute as a function of other variables of the datasets, and in the second phase, they apply previously designed model on the new and unseen datasets for determining the related class of each record. There are different methods for data classification such as Decision Trees (DT), Rule Based Methods, Logistic Regression (LogR), Linear Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Artificial Neural Networks (ANN), Linear Classifier (LC) and so on. We analyze three of the most common machine learning techniques, namely **decision tree**, **Support vector Machine** and **Apriori** algorithm. Decision tree, arpiori and SVM are using **iris.arff** dataset.

**Decision tree**: It is a predictive model which maps observations about an item to conclusions about the item's target value.(ID3,C4.5)

**Support vector Machine**: A promising new method for the classification of both linear and nonlinear data. In a nutshell, a support vector machine (SVM) is an algorithm that works as follows. It uses a nonlinear mapping to transform the original training data into a higher dimension. Within this new dimension, it searches for the linear optimal separating hyper plane.

**Apriori algorithm**: The Apriori Algorithm is an influential algorithm for mining frequent item-sets for boolean association rules. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time.

**Dataset:**

We used “iris.arff” dataset the basic information of iris is given below.

Relevant Information: This is perhaps the best known database to be found in the pattern recognition literature -

Predicted attribute: class of iris plant. This is an exceedingly simple domain.

Number of Instances: 150 (50 in each of three classes)

Number of Attributes: 4 numeric, predictive attributes and the class Attribute Information:

1. Sepal length in cm

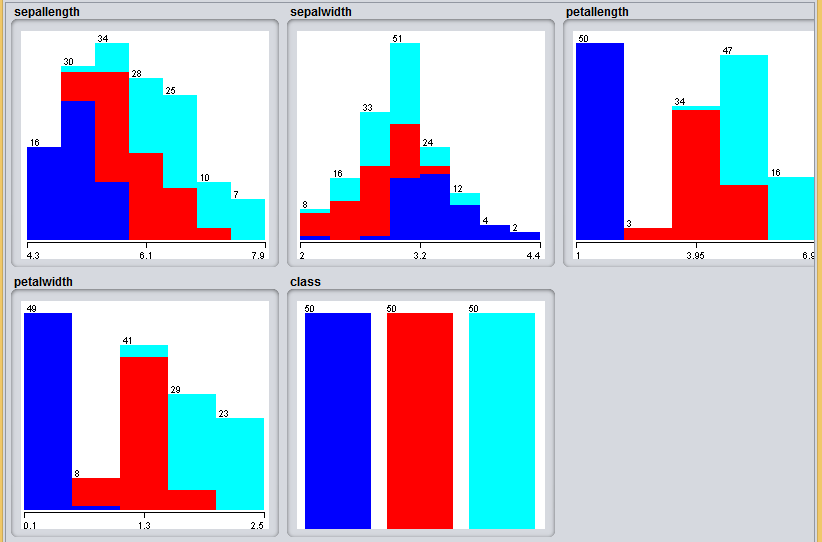
2. Sepal width in cm

3. Petal length in cm

4. Petal width in cm

5. Class: Iris Setosa , Iris Versicolour, Iris Virginica

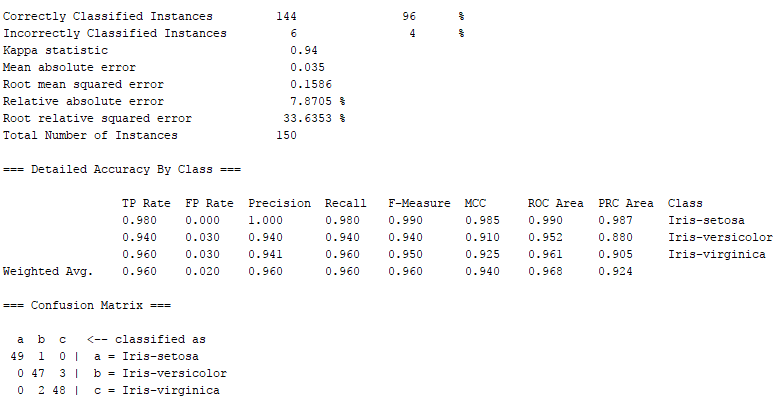
Missing Attribute Values: None.



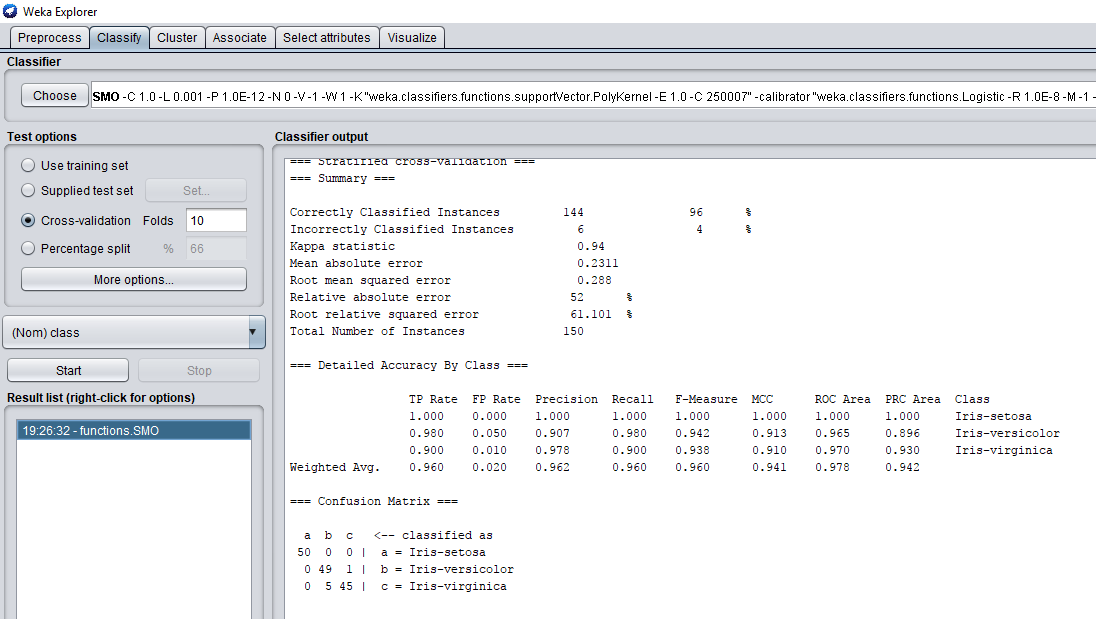
**Fig:** Attribute of Dataset

**Result:**

**J48:**



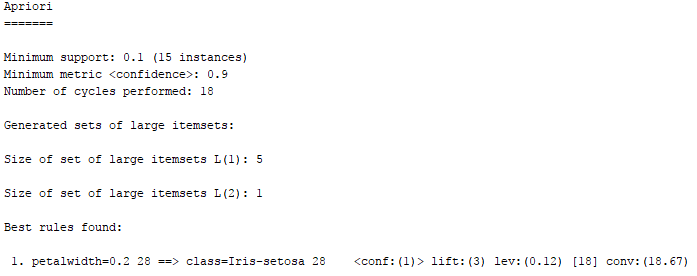
**SMO:**



**Comparison between decession tree and SVM:**

|  |  |  |
| --- | --- | --- |
| Error List | Decision tree | SVM |
| Kapa statistic | 0.94 | 0.94 |
| Mean absolute error | 0.035 | 0.2311 |
| Root mean squared error | 0.1586 | 0.288 |
| Relative absolute error | 7.8705% | 52% |
| Root relative squared error | 33.6353% | 61.101% |
| Total number of instance | 150 | 150 |

**Apriori:**



**Conclusion:**

In this paper, we compare four algorithm on iris dataset with some parameter. In iris dataset contain simple and class attribute. Decision tree will implement on iris dataset then it’s less efficient than all other like multilayer perceptron and Multiclass classifier. In all algorithm multilayer perceptron is more accurate.