International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 12 Ver. II || Dec 2018 || PP 63-65

Comparison of Prediction Success Performances for Classification Methods

Ayşe Eldem[1]

[1](*Department of Computer Engineering,KaramanoğluMehmetbey University, TURKEY)*

Abstract :*Methods such as obtaining meaningful data from databases, applying preprocesses to data, classification of data, estimating classes of new data using existing data are important issues in data mining. In the classification techniques, it is aimed to find the most appropriate class information according to the pre-defined class information of the data compared to other data mining techniques, clustering and association rules. A correctly trained model will provide more accurate classification of new data. In this study, many classification methods such as Decision Trees, Generalized Linear Model, Naive Bayes, Random Forest have been used in the classification of car dataset and performance comparison of these methods has been made.*

Keywords*-Classification, Classification Methods, Data Mining*

Date of Submission: 10-12-2018 Date of acceptance: 25-12-2018

I. Introduction

Classification is one of the methods of data mining. In the classification any data set is divided into train and test. Then, train the system with train data and test the system with the test data. The success of the system is determined by the accuracy of the test data never seen during the training phase. In this way, when a new data is received, it determines what kind of behavior.

Classification methods are used in many areas. Rotating machinery fault diagnosis were classified by using random forest, artificial neural network and support vector machine [1]. Landscape heterogeneity was classified by many methods such as Naive Bayes, Support Vector Machine, K-Nearest Neighbor, Bootstrap-aggregation ensemble of decision trees, Artificial Neural Network and Deep Neural Network [2]. The multi-label feature selection methods were described in detail [3].

In this study; classification methods are discussed. The car data set obtained from UCI [4] was used for classification by using different classification methods.

II. Material And Method

In this section; firstly, the data set used is mentioned. Then; Decision Trees, Generalized Linear Model, Naive Bayes, Random Forest classification methods used in this manuscript are described.

2.1.Dataset

The dataset is taken from UCI Machine Learning Repository [4]. The dataset has 1728 instances and 6 attributes. The car is classified some parametres like buying price, maintenance price, number of doors, capacity of person, luggage boot, safety of car. Each parametre has different values. The detailed information about parametres is available in Table 1. There are 4 class for the car classification which are called like unaccepted(unacc), accepted(acc), good(good) and very good(vgood).

Table 1.Attribute Values

Attribute Name	**Values**
buying	vhigh, high, med, low
maint	vhigh, high, med, low
doors	2, 3, 4, 5more
persons	2, 4, more
lug_boot	small, med, big
safety	low, med, high

2.2. Decision Trees

In supervised classification it is one of the most widely used method. For each attribute, entropy is calculated by gaining information. Then the root and leaves of the tree are determined according to the values found. ID3 algorithm was used in this study.

2.3. Generalized Linear Model

The model was developed by John Nelder and Robert Wedderburn [5]. A model which measures the relationship between attributes are called linear regression model. Generalized Linear Model is preferred for problem solving where dependent variables are continuous but not normally distributed [6].

2.4. Naive Bayes

It is the probabilistic classification method based on Bayes' theorem developed by Thomas Bayes. Naive Bayes classification method has many attributes and targets.

2.5. Random Forest

Random Forest was developed by Leo Breiman [7]. Using a combination of multiple decision tree is a model that allows to obtain the best classification. Random forest can be used in both classification and regression models. Overfitting can be avoided because it created different trees. In this study, 60 trees were formed.

III. Application

This study; was developed to measure the success of different classification models. Car dataset from UCI with 6 inputs was used in the study. It is classified as unaccepted, accepted, good and very good. The process steps are shown in Fig.1.Firstly dataset is divided for train and test.70% of the data set was used as train dataset and 30% as test dataset. Then, the system is trained with train dataset byDecision Trees, Generalized Linear Model, Naive Bayes, Random Forest algorithms. Decision Tree, Generalized Linear Model, Naive Bayes and Random Forest performance were showed in Table 1, Table 2, Table 3 and Table 4. Prediction and precision values were showed in tables. All the methods' performance comparison was detailed in Table 6. When the accuracy and runtime values were examined, Random Forest was the best for accuracy and Naive Bayes was the faster algorithm than the others also Naive Bayes was the worst algorithm for classification.

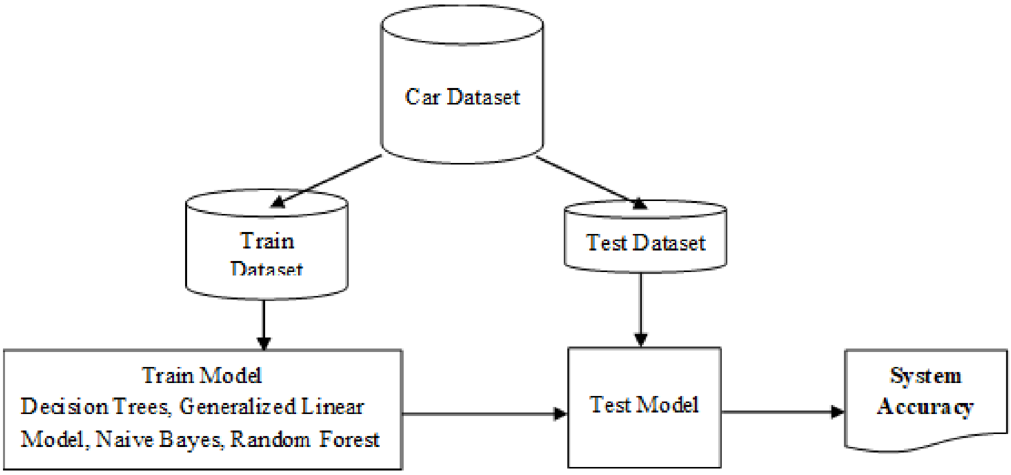

Fig. 2. Process Steps

Table 2. Decision Tree Performance

	True unacc	True acc	True vgood	True good	Class Precision
Predictionunacc	233	15	0	0	93.95%
Predictionacc	6	60	0	1	89.55%
Predictionvgood	0	2	13	3	72.22%
Predictiongood	3	0	0	10	76.92%
Class recall	96.28%	77.92%	100%	71.43%	

Table 3. Generalized Linear Model Performance

	True unacc	True acc	True vgood	True good	Class Precision
Predictionunacc	229	7	0	0	97.03%
Predictionacc	12	67	2	3	79.76%
Predictionvgood	0	1	11	0	91.67%
Predictiongood	1	2	0	11	78.57%
Class recall	94.63%	87.01%	84.62%	78.57%	

Table 4. Naive Bayes Performance

	True unacc	True acc	True vgood	True good	Class Precision
Predictionunacc	232	23	0	0	90.98%
Predictionacc	9	52	7	9	67.53%
Predictionvgood	0	0	6	0	100%
Predictiongood	1	2	0	5	62.50%
Class recall	95.87%	67.53%	46.15%	35.71%	

Table 5. Random Forest Performance

	True unacc	True acc	True vgood	True good	Class Precision
Predictionunacc	233	10	0	0	95.88%
Predictionacc	6	64	0	0	91.43%
Predictionvgood	0	2	11	0	84.62%
Predictiongood	3	1	2	14	70.00%
Class recall	96.28%	83.12%	84.62%	100%	

Table 6.Performance Comparison

Model	**Accuracy**	**Runtime**
DecisionTree	91.3%	171 ms
GeneralizedLinear Model	91.9%	3 s
NaiveBayes	85.3%	99 ms
RandomForest	93.1%	3 s

IV. Results

In this study, different classification methods have been applied by using car dataset. 30% of dataset is reserved as test data. According to the results obtained, Random Forest method has achieved the best classification accuracy. In addition, the fastest algorithm in terms of working time is Naive Bayes method.

Acknowledgements

This study was supported by KaramanoğluMehmetbey University.

References

[1]. T Han, D Jiang, Q Zhao, L Wang and K Yin, Comparison of random forest, artificial neural networks and support vector machine for intelligent diagnosis of rotating machinery,Transactions of the Institute of Measurement and Control, 40(8), 2018, 2681–2693.

[2]. S SHeydari and GMountraki, Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites,Remote Sensing of Environment, 204, 2018, 648-658.

[3]. R B Pereira, A Plastino, B Zadrozny and L H C Merschmann, Categorizing feature selection methods for multi-label classification,Artificial Intelligence Review, 49(1), 2018, 57-78

[4]. UCI Car Evaluation Dataset, https://archive.ics.uci.edu/ml/datasets/car+evaluation, Access Time: 10.08.2018

[5]. J A Nelder and R W M Wedderburn, Generalized Linear Models, J. Roy. Statist. Soc. A, 135, 1972, 370-384.

[6]. T Koç and M A Cengiz, Comparions of Estimation Methods in Generalized Linear Mixed Models with an Application, Karaelmas Science and Engineering Journal, 2 (2), 2012, 47-52.

[7]. L Breiman, Random Forests, Machine Learning, 45, 2001, 5–32

Ayşe Eldem" Comparison of Prediction Success Performances for Classification Methods"International Journal of Engineering Science Invention (IJESI), vol. 07, no. 12, 2018, 63-65