MET CS-699 DATA MINING FINAL PROJECT REPORT

Team member:

你的名字

Weilin Lu

DATA Resource URL：

***Introduction***

Data is an essential part of computer science, statistics, data analysis, and business analysis. More and more people are starting to work in the data analysis industry. Therefore, data mining is a must-learn technique for anyone who aspires to become a data analyst. Data mining refers to the process of searching for information hidden in a large amount of data through algorithms. Data mining is generally related to computer science and achieves these goals through several methods such as statistics, online analytical processing, intelligence retrieval, machine learning, expert systems (relying on past rules of thumb), and pattern recognition.

This data mining project finds whether there is a clear relationship between different data (purchase time, purchase quantity, indeed and customer gender) and the payment method chosen by the final customer through transaction data (excluding detailed transaction amount) of a website.

***Detailed description of the datasets***

Our data set is the historical purchase records of customers from a shopping website. A total of 26 columns and 500 rows of data. It includes purchase time, status, purchase quantity, price, discount, total, year, month, day, transaction number, name, age, address information and payment method.These data include numeric data and categorical data. All we have to do is to clean the useless data out of the original data, and then mine the correlation between the different data retained and the payment method chosen by the end customer.

***Tools to use***

The tool we use are Chrome and Weka. Chrome is a Google's browser. WEKA(Waikato Environment for Knowledge Analysis). It is an open source machine learning and data mining software based on JAVA environment.We use Chrome to find the data-set online and use Weka for data mining to get the confusion matrix

***5 Classification Algorithms***

***Naïve Bayes:***Bayesian classification algorithm is a classification method in statistics, which is a class of algorithms that use the knowledge of probability and statistics to classify.

***Logistic Regression:***It is a generalized linear regression analysis model, which belongs to supervised learning in machine learning. The model is trained by given n sets of data (training set), and after training is completed, the given set or sets of data (test set) are classified. Each set of data is composed of p indicators

***J48:***This algorithm is a kind of decision tree. Based on top-to-bottom strategy, recursive divide-and-conquer strategy, select an attribute to place at the root node, generate a branch for each possible attribute value, divide the instance into multiple subsets, each subset corresponds to a branch of the root node , then recursively repeat the process on each branch. Stop when all instances have the same classification.

***Random Forest:***Random Forest is a classifier that uses multiple trees to train and predict samples.

***Bagging:***The Bagging algorithm can be combined with other classification and regression algorithms to improve its accuracy and stability, and at the same time reduce the variance of the results to avoid over fitting.

Brief description of attribute selection methods we used and Set of attributes selected by each attribute selection method（简要说明我们使用的属性选择方法和每种属性选择方法选择的属性集合。）

Data mining Progress(Detailed description of data mining procedure (the procedure you actually followed) including all data preprocessing you performed./数据挖掘程序详细说明（你实际使用的程序）跟随）包括您执行的所有数据预处理。)

***Data mining result and evaluation***

We data-mined the raw data using five algorithms. After obtaining the confusion matrix of the original data, we trained the data based on five different Attributes and five different algorithms and finally obtained the accuracy and confusion matrix. From the final result comparison, RandomForest based on ClassifierAttributeEval has the highest data accuracy. Its accuracy rate is: 97.2727%

**Original Data-set（原始数据好像也需要分成66%训练和34%的测试来做。考完试以后注意一下。）**

***Naïve Bayes:***

Correctly Classified Instances: 90%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.964 | 0.181 | 0.871 | 0.964 | 0.915 | 0.800 | 0.947 | 0.960 | Paypal |
|  | 0.819 | 0.036 | 0.948 | 0.819 | 0.879 | 0.800 | 0.946 | 0.896 | Visa |
| Weighted average | 0.900 | 0.117 | 0.905 | 0.900 | 0.899 | 0.800 | 0.947 | 0.932 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 269 | 10 | a=paypal |
| 40 | 181 | b=Visa |

***Logistic***

Correctly Classified Instances: 91%

Detailed Accuracy By Class

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.918 | 0.100 | 0.921 | 0.918 | 0.919 | 0.818 | 0.938 | 0.945 | Paypal |
|  | 0.900 | 0.08 | 0.896 | 0.900 | 0.898 | 0.818 | 0.939 | 0.904 | Visa |
| Weighted average | 0.910 | 0.092 | 0.910 | 0.910 | 0.910 | 0.818 | 0.939 | 0.927 |  |

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 256 | 23 | a=paypal |
| 22 | 199 | b=Visa |

***J48***

Correctly Classified Instances: 94%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.982 | 0.113 | 0.916 | 0.982 | 0.948 | 0.880 | 0.981 | 0.982 | Paypal |
|  | 0.887 | 0.018 | 0.975 | 0.887 | 0.929 | 0.880 | 0.981 | 0.969 | Visa |
| Weighted average | 0.940 | 0.071 | 0.942 | 0.940 | 0.940 | 0.880 | 0.981 | 0.976 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 274 | 5 | a=paypal |
| 25 | 196 | b=Visa |

***Random Forest***

Correctly Classified Instances: 96.4%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.971 | 0.045 | 0.964 | 0.971 | 0.968 | 0.927 | 0.989 | 0.989 | Paypal |
|  | 0.955 | 0.029 | 0.963 | 0.955 | 0.959 | 0.927 | 0.989 | 0.989 | Visa |
| Weighted average | 0.9964 | 0.038 | 0.964 | 0.964 | 0.964 | 0.927 | 0.989 | 0.989 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 271 | 8 | a=paypal |
| 10 | 211 | b=Visa |

***Bagging***

Correctly Classified Instances: 94.8%

Detailed Accuracy By Class

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.953 | 0.059 | 0.953 | 0.953 | 0.953 | 0.895 | 0.989 | 0.991 | Paypal |
|  | 0.941 | 0.047 | 0.941 | 0.941 | 0.941 | 0.895 | 0.989 | 0.987 | Visa |
| Weighted average | 0.948 | 0.053 | 0.948 | 0.948 | 0.948 | 0.895 | 0.989 | 0.989 |  |

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 266 | 13 | a=paypal |
| 13 | 208 | b=Visa |

***The we split the data-set into two parts: 66% as training set, 33% as testing set***

***66% Traing:***

Attributes selected :

1. ***ClassifierAttributeEval:***
   1. Naïve Bayes

Correctly Classified Instances: 87.5758%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.973 | 0.247 | 0.833 | 0.973 | 0.879 | 0.757 | 0.945 | 0.958 | Paypal |
|  | 0.953 | 0.027 | 0.957 | 0.753 | 0.843 | 0.757 | 0.944 | 0.897 | Visa |
| Weighted average | 0.876 | 0.150 | 0.887 | 0.873 | 0.873 | 0.757 | 0.945 | 0.931 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 179 | 5 | a=paypal |
| 36 | 110 | b=Visa |

* 1. Logistics

Correctly Classified Instances: 87.5758%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.875 | 0.164 | 0.870 | 0.875 | 0.873 | 0.711 | 0.905 | 0.889 | Paypal |
|  | 0.836 | 0.125 | 0.841 | 0.836 | 0.838 | 0.711 | 0.903 | 0.872 | Visa |
| Weighted average | 0.858 | 0.147 | 0.857 | 0.858 | 0.858 | 0.711 | 0.904 | 0.882 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 161 | 23 | a=paypal |
| 24 | 122 | b=Visa |

* 1. J48

Correctly Classified Instances: 95.4545%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.978 | 0.075 | 0.942 | 0.978 | 0.960 | 0.908 | 0.984 | 0.978 | Paypal |
|  | 0.925 | 0.022 | 0.971 | 0.925 | 0.947 | 0.908 | 0.984 | 0.978 | Visa |
| Weighted average | 0.955 | 0.052 | 0.955 | 0.955 | 0.954 | 0.908 | 0.954 | 0.978 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 180 | 4 | a=paypal |
| 11 | 135 | b=Visa |

* 1. RandomForest:

Correctly Classified Instances: 97.2727%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.984 | 0.041 | 0.968 | 0.984 | 0.976 | 0.945 | 0.995 | 0.995 | Paypal |
|  | 0.959 | 0.016 | 0.979 | 0.959 | 0.969 | 0.945 | 0.995 | 0.995 | Visa |
| Weighted average | 0.973 | 0.303 | 0.973 | 0.973 | 0.973 | 0.945 | 0.995 | 0.995 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 181 | 3 | a=paypal |
| 6 | 140 | b=Visa |

* 1. Bagging:

Correctly Classified Instances: 95.1515%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.967 | 0.068 | 0.947 | 0.967 | 0.957 | 0.902 | 0.989 | 0.991 | Paypal |
|  | 0.932 | 0.033 | 0.958 | 0.932 | 0.944 | 0.902 | 0.989 | 0.989 | Visa |
| Weighted average | 0.952 | 0.053 | 0.952 | 0.952 | 0.951 | 0.902 | 0.989 | 0.990 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 178 | 6 | a=paypal |
| 10 | 136 | b=Visa |

1. ***RelifFAttributeEval***
   1. Naïve Bayes

Correctly Classified Instances: 86.3636%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.832 | 0.096 | 0.916 | 0.832 | 0.872 | 0.731 | 0.955 | 0.957 | Paypal |
|  | 0.904 | 0.168 | 0.810 | 0.904 | 0.854 | 0.731 | 0.955 | 0.957 | Visa |
| Weighted average | 0.864 | 0.128 | 0.869 | 0.864 | 0.864 | 0.731 | 0.955 | 0.957 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 153 | 31 | a=paypal |
| 14 | 132 | b=Visa |

* 1. Logistics

Correctly Classified Instances: 92.4242%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.929 | 0.082 | 0.934 | 0.929 | 0.932 | 0.847 | 0.951 | 0.964 | Paypal |
|  | 0.918 | 0.071 | 0.912 | 0.918 | 0.915 | 0.847 | 0.951 | 0.900 | Visa |
| Weighted average | 0.924 | 0.077 | 0.924 | 0.924 | 0.924 | 0.847 | 0.951 | 0.936 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 171 | 13 | a=paypal |
| 12 | 134 | b=Visa |

* 1. J48

Correctly Classified Instances: 93.9394%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.951 | 0.075 | 0.941 | 0.951 | 0.946 | 0.877 | 0.981 | 0.981 | Paypal |
|  | 0.925 | 0.049 | 0.938 | 0.925 | 0.931 | 0.877 | 0.981 | 0.971 | Visa |
| Weighted average | 0.939 | 0.064 | 0.939 | 0.939 | 0.939 | 0.877 | 0.981 | 0.976 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 175 | 9 | a=paypal |
| 11 | 135 | b=Visa |

* 1. Randomforest

Correctly Classified Instances: 96.0606%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.978 | 0.44 | 0.952 | 0.978 | 0.965 | 0.920 | 0.990 | 0.992 | Paypal |
|  | 0.938 | 0.022 | 0.972 | 0.938 | 0.955 | 0.920 | 0.990 | 0.983 | Visa |
| Weighted average | 0.961 | 0.062 | 0.961 | 0.961 | 0.961 | 0.920 | 0.990 | 0.988 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 180 | 4 | a=paypal |
| 9 | 137 | b=Visa |

* 1. Bagging

Correctly Classified Instances: 94.8485%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.957 | 0.062 | 0.951 | 0.957 | 0.954 | 0.896 | 0.988 | 0.990 | Paypal |
|  | 0.938 | 0.043 | 0.945 | 0.938 | 0.942 | 0.896 | 0.988 | 0.987 | Visa |
| Weighted average | 0.948 | 0.054 | 0.948 | 0.948 | 0.948 | 0.896 | 0.988 | 0.989 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 176 | 8 | a=paypal |
| 9 | 137 | b=Visa |

1. ***OneRAttributeEval***
   1. Naïve Bayes

Correctly Classified Instances: 90%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.902 | 0.103 | 0.917 | 0.902 | 0.910 | 0.798 | 0.954 | 0.953 | Paypal |
|  | 0.897 | 0.098 | 0.879 | 0.897 | 0.88 | 0.798 | 0.954 | 0.957 | Visa |
| Weighted average | 0.900 | 0.101 | 0.900 | 0.900 | 0.900 | 0.798 | 0.954 | 0.955 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 166 | 18 | a=paypal |
| 15 | 131 | b=Visa |

* 1. Logistics

Correctly Classified Instances: 92.4242%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.935 | 0.089 | 0.930 | 0.935 | 0.932 | 0.846 | 0.964 | 0.969 | Paypal |
|  | 0.911 | 0.065 | 0.917 | 0.911 | 0.914 | 0.846 | 0.964 | 0.959 | Visa |
| Weighted average | 0.924 | 0.078 | 0.924 | 0.924 | 0.924 | 0.846 | 0.964 | 0.964 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 172 | 12 | a=paypal |
| 13 | 133 | b=Visa |

* 1. J48

Correctly Classified Instances: 92.4242%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.935 | 0.089 | 0.930 | 0.935 | 0.932 | 0.846 | 0.964 | 0.969 | Paypal |
|  | 0.911 | 0.065 | 0.917 | 0.911 | 0.914 | 0.846 | 0.964 | 0.959 | Visa |
| Weighted average | 0.924 | 0.079 | 0.924 | 0.924 | 0.924 | 0.846 | 0.964 | 0.964 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 172 | 12 | a=paypal |
| 13 | 133 | b=Visa |

* 1. Randomforest

Correctly Classified Instances: 94.5455%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.957 | 0.068 | 0.946 | 0.957 | 0.951 | 0.889 | 0.987 | 0.990 | Paypal |
|  | 0.932 | 0.043 | 0.944 | 0.932 | 0.938 | 0.889 | 0.987 | 0.979 | Visa |
| Weighted average | 0.945 | 0.057 | 0.945 | 0.945 | 0.945 | 0.889 | 0.987 | 0.985 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 176 | 8 | a=paypal |
| 10 | 136 | b=Visa |

* 1. Bagging

Correctly Classified Instances: 94.5455%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.957 | 0.068 | 0.946 | 0.957 | 0.951 | 0.889 | 0.987 | 0.990 | Paypal |
|  | 0.932 | 0.043 | 0.944 | 0.932 | 0.938 | 0.889 | 0.987 | 0.979 | Visa |
| Weighted average | 0.945 | 0.057 | 0.945 | 0.945 | 0.945 | 0.889 | 0.987 | 0.985 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 176 | 8 | a=paypal |
| 10 | 136 | b=Visa |

1. ***CorrelationAttributeEval***
   1. Naïve Bayes

Correctly Classified Instances: 89.697%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.908 | 0.116 | 0.908 | 0.908 | 0.908 | 0.791 | 0.953 | 0.955 | Paypal |
|  | 0.887 | 0.092 | 0.884 | 0.884 | 0.884 | 0.791 | 0.953 | 0.956 | Visa |
| Weighted average | 0.897 | 0.106 | 0.897 | 0.897 | 0.897 | 0.791 | 0.953 | 0.955 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 167 | 17 | a=paypal |
| 17 | 129 | b=Visa |

* 1. Logistics

Correctly Classified Instances: 91.5152%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.918 | 0.089 | 0.929 | 0.918 | 0.923 | 0.828 | 0.935 | 0.959 | Paypal |
|  | 0.911 | 0.082 | 0.899 | 0.911 | 0.905 | 0.828 | 0.935 | 0.845 | Visa |
| Weighted average | 0.915 | 0.086 | 0.915 | 0.915 | 0.915 | 0.828 | 0.935 | 0.909 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 169 | 15 | a=paypal |
| 13 | 133 | b=Visa |

* 1. J48

Correctly Classified Instances: 94.2424%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.957 | 0.075 | 0.941 | 0..957 | 0.949 | 0.883 | 0.982 | 0.981 | Paypal |
|  | 0.925 | 0.043 | 0.944 | 0.925 | 0.934 | 0.883 | 0.982 | 0.972 | Visa |
| Weighted average | 0.942 | 0.061 | 0.942 | 0.9412 | 0.942 | 0.883 | 0.982 | 0.977 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 176 | 8 | a=paypal |
| 11 | 135 | b=Visa |

* 1. Randomforest

Correctly Classified Instances: 95.4545%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.957 | 0.048 | 0.962 | 0.957 | 0.959 | 0.908 | 0.990 | 0.993 | Paypal |
|  | 0.952 | 0.043 | 0.946 | 0.952 | 0.949 | 0.908 | 0.990 | 0.983 | Visa |
| Weighted average | 0.955 | 0.9046 | 0.955 | 0.955 | 0.955 | 0.908 | 0.990 | 0.989 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 176 | 8 | a=paypal |
| 7 | 139 | b=Visa |

* 1. Bagging

Correctly Classified Instances: 94.8485%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.957 | 0.062 | 0.951 | 0.957 | 0.954 | 0.896 | 0.988 | 0.990 | Paypal |
|  | 0.938 | 0.043 | 0.945 | 0.938 | 0.942 | 0.896 | 0.988 | 0.987 | Visa |
| Weighted average | 0.948 | 0.054 | 0.948 | 0.948 | 0.948 | 0.896 | 0.988 | 0.989 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 176 | 8 | a=paypal |
| 9 | 137 | b=Visa |

1. ***GainRatioAttributeEval***
   1. Naïve Bayes

Correctly Classified Instances: 83.0303%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.902 | 0.260 | 0.841 | 0.902 | 0.856 | 0.656 | 0.912 | 0.926 | Paypal |
|  | 0.740 | 0.098 | 0.857 | 0.740 | 0.794 | 0.656 | 0.912 | 0.899 | Visa |
| Weighted average | 0.830 | 0.188 | 0.833 | 0.830 | 0.828 | 0.656 | 0.912 | 0.914 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 166 | 18 | a=paypal |
| 38 | 108 | b=Visa |

* 1. Logistics

Correctly Classified Instances: 86.0606%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.870 | 0.151 | 0.879 | 0.870 | 0.874 | 0.718 | 0.924 | 0.939 | Paypal |
|  | 0.849 | 0.130 | 0.838 | 0.849 | 0.844 | 0.718 | 0.924 | 0.855 | Visa |
| Weighted average | 0.861 | 0.142 | 0.861 | 0.861 | 0.861 | 0.718 | 0.924 | 0.902 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 160 | 24 | a=paypal |
| 22 | 124 | b=Visa |

* 1. J48

Correctly Classified Instances: 91.2121%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.918 | 0.096 | 0.923 | 0.918 | 0.921 | 0.822 | 0.967 | 0.966 | Paypal |
|  | 0.904 | 0.082 | 0.898 | 0.904 | 0.901 | 0.822 | 0.967 | 0.954 | Visa |
| Weighted average | 0.912 | 0.090 | 0.912 | 0.912 | 0.912 | 0.822 | 0.967 | 0.961 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 169 | 15 | a=paypal |
| 14 | 132 | b=Visa |

* 1. Randomforest

Correctly Classified Instances: 93.6364%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.957 | 0.089 | 0.931 | 0.957 | 0.944 | 0.871 | 0.985 | 0.988 | Paypal |
|  | 0.911 | 0.043 | 0.943 | 0.911 | 0.927 | 0.871 | 0.985 | 0.983 | Visa |
| Weighted average | 0.936 | 0.069 | 0.937 | 0.936 | 0.936 | 0.871 | 0.985 | 0.986 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 176 | 8 | a=paypal |
| 13 | 133 | b=Visa |

* 1. Bagging

Correctly Classified Instances: 90.9091%

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | FP | Precision | Recall | F-Measure | MCC | ROC Area | PRC Area | class |
|  | 0.929 | 0.116 | 0.910 | 0.929 | 0.919 | 0.815 | 0.974 | 0.975 | Paypal |
|  | 0.884 | 0.071 | 0.908 | 0.884 | 0.896 | 0.815 | 0.974 | 0.972 | Visa |
| Weighted average | 0.909 | 0.096 | 0.909 | 0.909 | 0.909 | 0.815 | 0.974 | 0.975 |  |

Detailed Accuracy By Class

Confusion Matrix

|  |  |  |
| --- | --- | --- |
| a | b | Classified as |
| 171 | 13 | a=paypal |
| 17 | 129 | b=Visa |

Conclusion

Summing the above data and accuracy, it can be found that Random Forest based on ClassifierAttributeEval has the highest data accuracy. We can also conclude that even with the same Attribute, using different algorithms can have a huge impact on the results. Therefore, when mining data, it is necessary to conduct multiple experiments to ensure that the best data model is obtained, so as to obtain the most correct data results.