Author – Ranjith Dhanaraj.

Comparison of Filter and Wrapper Functions using Weka

Filter – Feature Selection was performed by analyzing the accuracy of the prediction model and selecting the least number of attributes that provide us with the most accuracy. The attributes were ranked using **Information Gain filter**. There are 5 ways to select the features –

1. Select the top K features.
2. Select top 50%.
3. Select Features with Information Gain > 50%.
4. Subset of features with non-zero IG.
5. Evaluate classification performance using feature subsets of increasing size. Starting with feature with highest Information Gain, then add the next feature. Measure the accuracy for each subset using cross-validation. Choose the final subset giving the highest accuracy.

```
Ranked attributes:
 0.0211783    6 handsetAge
 0.0205821   25 lifeTime
 0.0069636   13 avgMins
 0.0066154   17 callMinutesChangePct
 0.0058219   14 avgrecurringCharge
 0.0051004    7 smartPhone
 0.0031704    9 creditRating
 0.0031557   27 numRetentionCalls
 0.0026841   20 avgOutCalls
 0.0026106   26 lastMonthCustomerCareCalls
 0.0024857   21 avgInCalls
 0.0020887   19 avgReceivedMins
 0.0018673   12 avgBill
 0.0014273   24 avgDroppedCalls
 0.0012043   28 numRetentionOffersAccepted
 0.0004432    2 marriageStatus
 0.000139    11 creditCard
 0.0001228   10 homeOwner
 0.0000893    3 children
 0           29 newFrequentNumbers
 0            5 numHandsets
 0            4 income
 0           15 avgOverBundleMins
 0            8 currentHandsetPrice
 0           16 avgRoamCalls
 0           23 peakOffPeakRatioChangePct
 0           22 peakOffPeakRatio
 0           18 billAmountChangePct
 0            1 age
```

The 5th option seems to be the best method since

- Option 1 has no reliable way of selecting K.
- Option 2 would select too many attributes which would induce high dimensionality and induce more noise.
- Option 3 would select none of the attributes.
- Option 4 again will have too many attributes thereby inducing more noise and high dimensionality. High Dimensionality is a problem since it increases computation cost.
- Option 5 as this will select the smallest amount of features before accuracy suffers and as the training set is quite large, we do not need to worry about overfitting with this method.

As we can see the attributes with the most score are – handsetAge, lifeTime, avgMins, callMinutesChangePct, avgrecurringCharge and smartPhone. Selection of the attributes were done in descending order of IG score. Each attribute was selected and classification was performed. The subset of attributes which had the highest accuracy were finally selected. The aforementioned attributes were used to train the prediction model with Naïve Bayes and 10-NN classifier. The value of 10 was selected for KNN because the accuracy was the highest and remained stagnant later on while maintaining relatively low validation errors. 10-Fold Cross validation was performed while assessing the accuracy of the prediction models.

Classifier	Accuracy
10-NN	57.7158%
Naïve Bayes	56.2947%

Wrapper – Feature selection was performed using Wrapper Subset Evaluation with Greedy Stepwise Forward Selection. Classifier in the wrapper was selected to be 10-NN model as well. The selection was done again through trial and error. Features selected using Naïve Bayes as the Wrapper Function Classifier were less accurate when compared to KNN. Since the accuracy of the system was the best at 10-NN in the previous methodology, the value of K was again set to be 10 in the wrapper evaluation function Classifier. Once the features were obtained, classification with 10-fold Cross Verification was performed with 10-NN Classifier and Naïve Bayes Classifier. We can observe that the accuracy using 10-NN is higher compared to Naïve Bayes.

Classifier	Accuracy

```
Selected attributes: 7,9,25,27 : 4
                     smartPhone
                     creditRating
                     lifeTime
                     numRetentionCalls
```

Subset Selected with Filter - handsetAge, lifeTime, avgMins, callMinutesChangePct, avgrecurringCharge and smartPhone.

Subset Selected with Wrapper – smartPhone, creditRating, lifeTime, numRetentionCalls.

Information Gain filter tends to extract attributes with the most amount of statistical importance. It is based on a mathematical evaluation function. Filter methods consider correlation or mutual information present within the features. Wrapper method on the other hand finds features by measuring the "goodness" of subsets of features while training the model and evaluates which features are best suited for the dataset. A classification algorithm can be chosen and hence the algorithm's performance is used to perform the evaluation. Filters return just a ranking of all the attributes of the dataset whereas a Wrapper method returns definite attributes which can be used for maximum efficiency.

We can observe that the attributes smartPhone and lifeTime are present in both the feature subsets. Filter methods have more attributes in the given example and this might again be attributed to the fact that evaluation occurs on a mathematical basis and not on features that best describe a dataset. Wrapper methods are more computationally expensive when compared to Filter methods but produce models which are more accurate in general. We can use a Filter method to get a prototype for a feasibility check on a dataset. Once we do know that we can achieve our intended goal we can use Wrapper methods to improve and fine tune the model based on our needs. But in our case the accuracy of the Filter method is higher when compared to Wrapper method. This maybe because of the fact that we use more attributes and that two of those attributes have more correlation to other attributes in the dataset.

Evaluation – From the tables below we can see that in general the accuracy of the Wrapper mode is lesser compared to that of Filter methods. The combination with most accuracy is a Filter method with 10-NN. This combination has an accuracy rate of 57.71% compared to its next closest competitor which is a wrapper method with 10-NN with accuracy of 57%. The Filter method includes more attributes but at the same time has a higher accuracy rate but lower ROC Area of 0.597 to the wrapper method which has a ROC area of 0.598. The answer to the question which combination to use gets quite muddy since the accuracy of the Filter method is higher but the ROC area is lower with respect to Wrapper-10-NN combination.

Preprossessing Method	Classifier	Accuracy	ROC Area	Attributes	Confusion Matrix
Filter	10-NN	57.7158%	0.597	1. handsetAge 2. lifetime 3. avgMins 4. callMinutesChangePct 5. avgrecurringCharge 6. smartPhone	a b <-- classified as 3168 1575 | a = true 2442 2315 | b = false
	Naïve Bayes	56.2947%	0.581		a b <-- classified as 2555 2188 | a = true 1964 2793 | b = false
Wrapper	10-NN	57%	0.598	1. smartphone 2. creditRating 3. lifetime 4. numRetentionCalls.	a b <-- classified as 2910 1833 | a = true 2252 2505 | b = false
	Naïve Bayes	54.6632%	0.569		a b <-- classified as 1694 3049 | a = true 1258 3499 | b = false