DATA MINING

CSE 454

HOMEWORK 4

CAN BEYAZNAR

161044038

1) Which technique has given better results in terms of f1 score? (filter feature selection or wrapper feature selection) Was it expected?

The Filter methodology uses the selected metric to identify irrelevant attributes and also filter out redundant columns from your models. It gives you the option of isolating selected measures that enrich your model. The columns are ranked following the calculation of the feature scores.

By choosing and implementing the right features, you can potentially improve the accuracy and efficiency of your classification models.

The Wrapper methodology considers the selection of feature sets as a search problem, where different combinations are prepared, evaluated and compared to other combinations. A predictive model is used to evaluate a combination of features and assign model performance scores.

The main differences between the filter and wrapper methods for feature selection are:

- Filter methods measure the relevance of features by their correlation with dependent variable while wrapper methods measure the usefulness of a subset of feature by actually training a model on it.
- Filter methods are much faster compared to wrapper methods as they do not involve training the models. On the other hand, wrapper methods are computationally very expensive as well.
- Filter methods use statistical methods for evaluation of a subset of features while wrapper methods use cross validation.
- Filter methods might fail to find the best subset of features in many occasions but wrapper methods can always provide the best subset of features.
- Using the subset of features from the wrapper methods make the model more prone to overfitting as compared to using subset of features from the filter methods.

In Filter feature selection methods, it works faster because there is no training process like in the wrapper. However, it is not expected to perform as well as the wrapper method. Because Wrapper uses a model, it tries to get the best solution. However, the filter doesn't do that. Wrapper is more expensive than filter methods. However, better results are expected.

```
Filter test
Correlation results :
[0.22189815303398835, 0.4665813983068757, 0.06506835955033308, 0.07475223191831948, 0.1305479548840481, 0.29269466264444666, 0.17384406565296076, 0.2383559830271975]


Accuracy :  94.47916666666667  Precision :  0.618421052631579  Recall :  0.6619718309859155  F1 :  0.6394557823129252
Accuracy :  88.50260416666667  Precision :  0.5751633986928104  Recall :  0.6984126984126984  F1 :  0.6308243727598566

Wrapper test
Accuracy :  89.4921875  Precision :  0.6267605633802817  Recall :  0.6267605633802817  F1 :  0.6267605633802817
Accuracy :  91.51041666666667  Precision :  0.5947712418300654  Recall :  0.7222222222222222  F1 :  0.6523297491039427
```

Judging by the results, the wrapper gets a higher result from the filter as expected.

2) Which technique has given better results in terms of f1 score? (PCA or LDA)? Was it expected?

PCA is an unsupervised machine learning method that is used for dimensionality reduction. The main idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of many variables correlated with each other, either heavily or lightly, while retaining the variation present in the dataset, up to the maximum extent.

LDA is a supervised machine learning method that is used to separate two groups/classes. The main idea of linear discriminant analysis(LDA) is to maximize the separability between the two groups so that we can make the best decision to classify them. **LDA is like PCA which helps in dimensionality reduction, but it focuses on maximizing the separability among known categories by creating a new linear axis and projecting the data points on that axis.** LDA doesn't work on finding the principal component, it basically looks at what type of point/features/subspace gives more discrimination to separate the data.

Both LDA and PCA are linear transformation techniques: LDA is a supervised whereas PCA is unsupervised – PCA ignores class labels. In contrast to PCA, LDA attempts to find a feature subspace that maximizes class separability.

When I compare the LDA and PCA methods I have applied, I get a better result than PCA. I had problems while transferring the naive bayes classification I implemented to LDA and PCA outputs due to the problems caused by my dataset. So I am not sure if the printouts are correct. The outputs are as follows.

```
-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-Part4 TEST-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-
-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-
PCA Test
F1 Scores: [0.9882352941176471, 0.15294117647058825, 0.8588235294117647, 0.8941176470588236, 0.32941176470588235, 0.9176470588235294, 0.8470588235294118, 0.9411764705882353, 0.9647058823529412]
Mean F1: 76.601%




-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-
-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-Part5 TEST-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-
-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-
LDA Test
F1 Scores: [0.3411764705882353, 0.5764705882352941, 0.6352941176470588, 0.6823529411764706, 0.6470588235294118, 0.6235294117647059, 0.6588235294117647, 0.6823529411764706, 0.6941176470588235]
Mean F1: 61.569%
```

3) Have the filter feature selection and wrapper feature selection technique given similar set of features? Which attributes are different?

When I examined the outputs I obtained, I saw that the values obtained in the filter were different from the values obtained in the wrapper. It has been observed that some values with low values in the filter are high in the wrapper.

```
Correlation results :
[0.22189815303398835, 0.4665813983068757, 0.06506835955033308, 0.07475223191831948, 0.1305479548840481, 0.29269466264444666, 0.17384406565296076, 0.2383559830271975]
```

```
wrapper
[[2.56929998][0.69564534][0.85589665][1.12762151][1.05199216][1.9260283][2.20545605][1.56772616]]
```

4) Which technique has given better results? (feature selection or dimention reduction)? Was it expected?

When the outputs are analyzed, the highest f1 is seen in the PCA method. However, there was not much difference in results between the methods. Some methods give incomplete or low results for various reasons. The order of the F1 results is as follows.

1) **PCA**
2) **Naive-Bayes**
3) **Wrapper**
4) **Filter**
5) **LDA**

```
Naive-Bayes test
Accuracy :  92.515625  Precision :  0.6764705882352942  Recall :  0.647887323943662  F1 :  0.6618705035971224
Accuracy :  90.51302083333333  Precision :  0.5960264900662252  Recall :  0.7142857142857143  F1 :  0.6498194945848376

Filter test
Correlation results :
[0.22189815303398835, 0.4665813983068757, 0.06506835955033308, 0.07475223191831948, 0.1305479548840481, 0.29269466264444666, 0.17384406565296076, 0.2383559830271975]


Accuracy :  94.47916666666667  Precision :  0.618421052631579  Recall :  0.6619718309859155  F1 :  0.6394557823129252
Accuracy :  88.50260416666667  Precision :  0.5751633986928104  Recall :  0.6984126984126984  F1 :  0.6308243727598566

Wrapper test
Accuracy :  89.4921875  Precision :  0.6267605633802817  Recall :  0.6267605633802817  F1 :  0.6267605633802817
Accuracy :  91.51041666666667  Precision :  0.5947712418300654  Recall :  0.7222222222222222  F1 :  0.6523297491039427

PCA Test
F1 Scores: [0.9882352941176471, 0.15294117647058825, 0.8588235294117647, 0.8941176470588236, 0.32941176470588235, 0.9176470588235294, 0.8470588235294118, 0.9411764705882353, 0.9647058823529412]
Mean F1: 76.601%

LDA Test
F1 Scores: [0.3411764705882353, 0.5764705882352941, 0.6352941176470588, 0.6823529411764706, 0.6470588235294118, 0.6235294117647059, 0.6588235294117647, 0.6823529411764706, 0.6941176470588235]
Mean F1: 61.569%
```


1- https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables

2- https://www.explorium.ai/blog/demystifying-feature-selection-filter-vs-wrapper-methods/

3- https://medium.com/analytics-vidhya/pca-vs-lda-vs-t-sne-lets-understand-the-difference-between-them-22fa6b9be9d0

4- https://sebastianraschka.com/faq/docs/lda-vs-pca.html