

## CHAPTER 7

### DESIGN AND DEVELOPMENT OF HYBRID CLASSIFIERS FOR SENTIMENT ANALYSIS

#### 7.1 INTRODUCTION

Ethem Alpaydin (2010) described that multiple base-learners can be combined to generate the final output. There are two methods of combining the base-learners. They are:

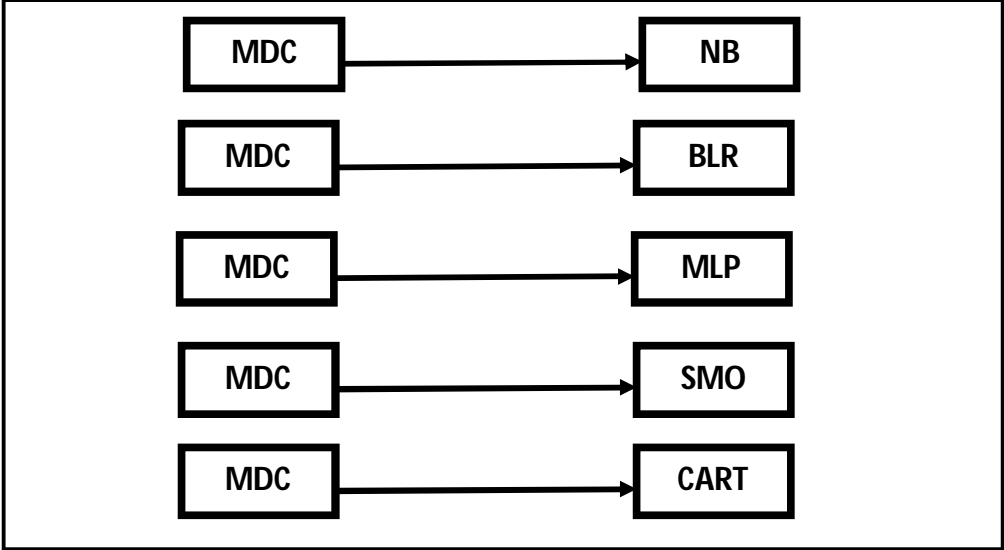
- **Multi-expert combination** – In this method, the base-learners work in parallel to generate the final output. This method is further divided into two categories viz. 1. Global approach 2. Local approach. In the *global approach* all the base-learners generate an output and all these outputs are used to determine the final output. In the *local approach*, based on the input, one or few of the base-learners are selected for determining the final output.
- **Multi-stage combination** – In this method, the base-learners are used in series. The next base-learner is trained with or tested on only the instances where the previous classifiers are not confident.

Rudy Prabowo and Mike Thelwall (2009) demonstrated the multi-stage combination of classifiers result in enhanced classification accuracy.

They used, rule-based classifier (RBC), statistics-based classifier (SBC), general inquirer-based classifier (GIBC) and SVM in some sequence to achieve better effectiveness.

**7.2 HYBRID OF MDC AND ML BASED CLASSIFIERS**

Inline with the concepts discussed in the section 7.1, the wrongly classified documents by MDC were classified by the ML based classifiers like NB, BLR, MLP, SMO and CART using WEKA 3.6.3. The hybrid classification was carried out in two stages as shown in Figure 7.1. For the investigations on the effect of combining multiple classifiers, the dataset LDS25000 has been used.



**Figure 7.1 The hybrid classification experimental procedure for MDC**

The MDC misclassified 7281 documents of LDS25000. In the 7281 documents, 3370 are negative documents misclassified as positive (FP) and 3911 positive documents misclassified as negative (FN). The feature vectors of these misclassified documents were separated out from the RTDM of LDS25000 and given as input to ML classifiers. Table 7.1 shows the

performance of various ML classifiers when used as a hybrid after the first-stage classification done by MDC. The performance reported in Table 7.1 is based on a ten-fold cross validation test.

**Table 7.1 Performance of various ML classifiers on the documents misclassified by MDC**

<b>Classifier</b>	<b>P</b>	<b>R</b>	<b>F-Measure</b>	<b>Accuracy</b>
NB	0.94	0.73	0.82	0.81
BLR	0.89	0.85	0.87	0.88
MLP	0.95	0.96	0.96	0.96
SMO	0.89	0.84	0.87	0.88
CART	0.95	0.94	0.95	0.95

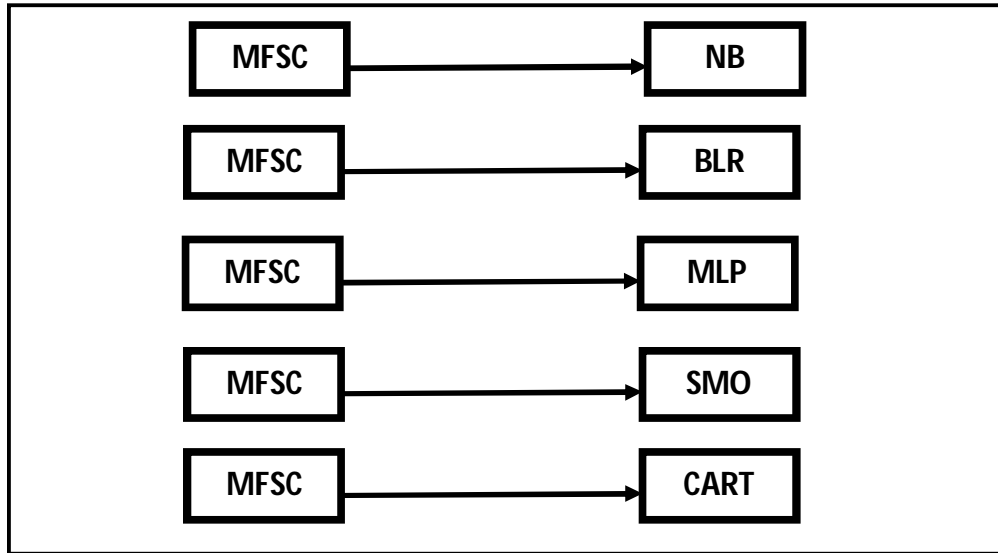
Total reviews -7281 (3911 positive and 3370 negative)

MLP and CART classified the documents with 95.92% and 95% respectively. The precision, recall and F-Measure score of MLP is comparatively better than CART.

Hence, the performance of MLP has been considered for discussion. Of the 3370 negative documents, 3189 have been correctly classified and of the 3911 positive documents, 3791 have been correctly classified. Thus a total of 24699 documents of LDS25000 have been correctly classified by the hybrid of MDC and MLP classifier. This amounts to an accuracy of 98.8%.

### **7.3 HYBRID OF MFSC AND ML BASED CLASSIFIERS**

Inline with the concepts discussed in the section 7.1, the wrongly classified documents by MFSC were classified by the ML based classifiers like NB, BLR, MLP, SMO and CART using WEKA 3.6.3. Only in two stages the hybrid classification was carried out as shown in Figure 7.2.



**Figure 7.2 The hybrid classification experimental procedure for MFSC**

The MFSC misclassified 7644 documents of LDS25000. In the 7644 documents, 3503 are negative documents misclassified as positive (FP) and 4141 positive documents misclassified as negative (FN). The feature vectors of these misclassified documents were separated out from the RTDM-FS of LDS25000 and given as input to ML classifiers. Table 7.2 shows the performance of various ML classifiers when used as a hybrid after the first-stage classification done by MFSC. The performance reported in Table 7.2 is based on a ten-fold cross validation test.

**Table 7.2 Performance of various ML classifiers on the documents misclassified by MFSC**

<b>Classifier</b>	<b>P</b>	<b>R</b>	<b>F-Measure</b>	<b>Accuracy</b>
NB	0.75	0.84	0.79	0.76
BLR	0.87	0.94	0.90	0.89
MLP	0.94	0.99	0.97	0.96
SMO	0.94	0.94	0.94	0.94
CART	0.99	0.99	0.99	0.99

Total reviews 7644 (4141 positive and 3503 negative)

MLP and CART classified the documents with 96.33% and 99.28% respectively. The precision, recall and F-Measure score of CART is comparatively better than MLP.

Hence the performance of CART has been considered for discussion. Of the 3503 negative documents, 3482 have been correctly classified and of the 4141 positive documents, 4107 have been correctly classified. Thus a total of 24945 documents of LDS25000 have been correctly classified by the hybrid of MFSC and CART. This amounts to an accuracy of 99.78%.

#### 7.4 HYBRID OF FLC AND ML BASED CLASSIFIERS

Inline with the concepts discussed in the section 7.1, the wrongly classified documents by FLC were classified by the ML based classifiers like NB, BLR, MLP, SMO and CART using WEKA 3.6.3. Only in two stages the hybrid classification was carried out as shown in Figure 7.3

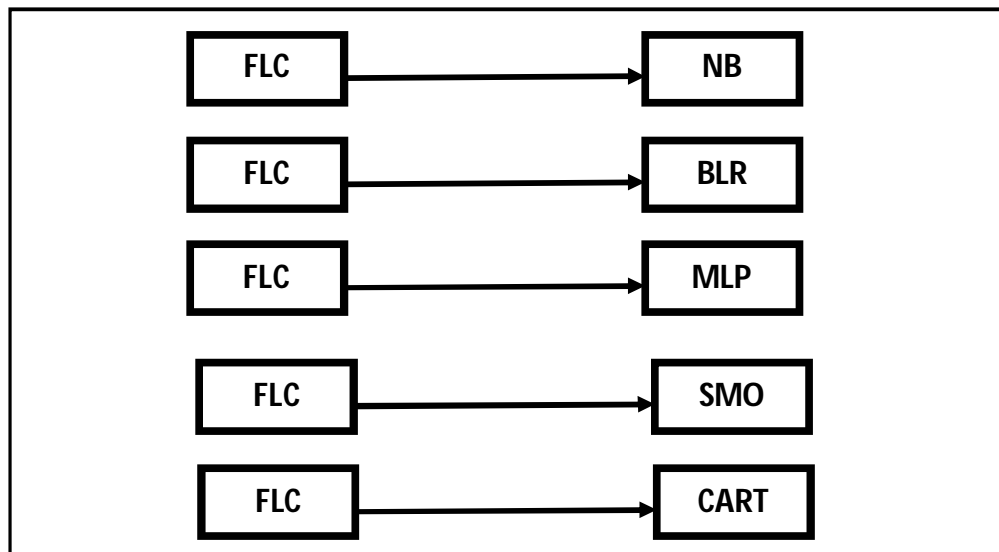


Figure 7.3 The hybrid classification experimental procedure for FLC

The FLC misclassified 8544 documents of LDS25000. In the 8544 documents, 2082 are negative documents misclassified as positive (FP) and 6462 positive documents misclassified as negative (FN). The feature vectors of these misclassified documents were separated out from the RTDM-FS of LDS25000 and given as input to ML classifiers. Table 7.3 shows the performance of various ML classifiers when used as a hybrid after the first-stage classification done by FLC. The performance reported in Table 7.3 is based on a ten-fold cross validation test.

**Table 7.3 Performance of various ML classifiers on the documents misclassified by FLC**

<b>Classifier</b>	<b>P</b>	<b>R</b>	<b>F-Measure</b>	<b>Accuracy</b>
NB	0.91	0.93	0.92	0.88
BLR	0.87	0.99	0.93	0.88
MLP	0.97	0.98	0.97	0.96
SMO	0.92	0.98	0.95	0.93
CART	0.97	0.98	0.98	0.97

Total reviews 8544 (6462 positive and 2082 negative)

MLP and CART classified the documents with 95.99% and 96.66% respectively. The precision, recall and F-Measure score of CART is comparatively better than MLP. Hence, the performance of CART has been considered for discussion. Of the 2082 negative documents, 1916 have been correctly classified and of the 6462 positive documents, 6343 have been correctly classified. Thus a total of 24715 documents of LDS25000 have been correctly classified by the hybrid of FLC and CART. This amounts to an accuracy of 98.9%.

## 7.5 DISCUSSION AND INFERENCE

Table 7.4 shows the overall classification performance of the best performing hybrid combination of MDC, MFSC and FLC for LDS25000 dataset.

**Table 7.4 Best performing hybrid classifiers (for LDS 25000)**

Hybrid classifier	P	R	F-Measure	A
MDC+MLP	0.986	0.99	0.988	98.8%
MFSC+CART	0.998	0.997	0.997	99.78%
FLC+CART	0.987	0.99	0.988	98.9%

As individual classifier, the order of MDC, MFSC and FLC in terms of classification performance is:

**MDC (70.8%) > MFSC (69%) > FLC (66%)** (Based on the performance on LDS25000)

As hybrid classifiers, the order of MDC, MFSC and FLC in terms of classification performance is:

**MFSC (99.78%) > FLC (98.9%) > MDC (98.8%)** (Based on the performance on LDS25000 using the multi-stage combination approach)

- The higher performance of MFSC as a hybrid is due to the excellent classification performance of CART at the second stage. CART performed with 99.28% of accuracy on classifying the documents that were misclassified by the MFSC.
- The FLC has moderate classification accuracy (66%) as stand-alone classifier. But as a hybrid with CART, the accuracy jumps to 96.66%, which is as good as the performance of MDC→MLP hybrid and MFSC→CART hybrid.