# *Chapter 8. Training and Testing Time*

In recent years, a robust, rapid and accurate email filtering system is becoming a primary and important need of the organisations and the Internet Service Providers. The researchers and developers always strive for experimenting and designing such filtering system that would be capable of providing excellent accuracy, sensitivity towards accuracy with low false positive rate, rapid training and real time fast testing. Several Machine Leaning Classifiers and different studies have been experimented in this area, but none of them was perfect due to some trade-off they have with each other. In the above chapters, many studies have been performed to capture the best classifier in terms of classification accuracy and false positive rate. Amongst all, best classifiers have been picked for further evaluation where the credibility of the selected classifiers has been analysed with all the performance metrics including training and testing time.

In this study, Enron email corpus is selected for observing the actual classification mechanisms where the training and testing sets are different. The corpuses taken in the above studies have not been included in this study because selected classifiers have already been tested on them. For building a robust model, training is done with 6000 most complex files of Enron corpus and then the entire (31510 email files) Enron corpus is tested on it.

## 8.1 Evaluation of Robust Email Filtering Models:

Experimental design and the description of this study have been given below:

### *8.1.1 Aim of this study:*

A comparative study is done to evaluate the excellent classification models on the basis of all the metrics including training and testing time.

### *8.1.2 Corpora for this study:*

All versions of the Enron Email corpus has been preferred for this study where training of the classifiers is done with the most complex Enron version (5, 6) with 6000 email files for generating various classification models and further, all email files including all versions (31510 email files) have been tested on the filtering models.

### *8.1.3 Feature Subset Search Techniques:*

Greedy Subset Feature search method is taken for this study.

### *8.1.4 Classifier for this study:*

Bayesian, Boosted Bayesian (*BB*), Support Vector Machine (*SVM*), and Enhanced Genetic Programming (*EGP*) classifiers are used for this study.

### *8.1.5 System Design:*

After Pre-processing (feature extraction and feature search), 38 most informative features for Enron (version 5, 6) and Enron (all versions) corpus are selected. Thereafter, 6000 email files of Enron (version 5, 6) are incorporated into training, and entire 31510 email files of Enron (all versions) are tested on different classification models.

### *8.1.6 Evaluation Metrics:*

Classification accuracy, F-Value, False Positive Rate, Training Time, and Testing Time are the evaluation metrics of this study.

### *8.1.7 Results and Analysis:*

In this study, the classifiers selected from the previous studies are trained on the 6000 email files of Enron (Version 5, 6) with the help of 38 features selected by the Greedy Stepwise method. For ensemble, different number of weak Bayesian and weak Genetic Programming (*GP*) classifiers are used to construct Boosted Bayesian (*BB*) and Enhanced Genetic Programming (*EGP*) classifiers that are further compared together with Bayesian and Support Vector Machine (*SVM*). This study is categorised into five parts where the first part analyses the results of classification accuracy and F-Value. Second part observes the classifiers' performance on the basis of False Positive Rate. The Third and Fourth parts will analyse the credibility of the classifiers in terms of Training and Testing Time respectively, and the last part finally observes the excellent classification model when analysis is done with all metrics.

*8.1.7.1 Analysis with Classification Accuracy and F-Value:*

The results of the classification accuracy and F-value have been demonstrated in Tables 44, 45 and Figures 57, 58 where the iterations actually indicate the number of classifiers ensemble for BB and EG. Iterations have not been performed for Bayesian and SVM. Some of observations are identified from the results that are shown below:

Table 44. Percentage Accuracy for Machine Learning Classifiers

| Percentage Accuracy | | | | |
|---|---|---|---|---|
| W/O Iterations | | With Iterations | | |
| Bayesian | SVM | Iterations | BB | EGP |
| 85.1 | 86.4 | 1 | 85 | 80.5 |
| | | 5 | 85.2 | 82.3 |
| | | 10 | 85.2 | 85.7 |
| | | 15 | 85.2 | 85.2 |
| | | 20 | 85.2 | 85.1 |
| | | 25 | 85.2 | **86.5** |
| | | 30 | 85.2 | 86.2 |
| | | 35 | 85.2 | 86.3 |
| | | 40 | 85.2 | **86.4** |

Table 45. Percentage F-Value for Machine Learning Classifiers

| Percentage F-Value | | | | |
|---|---|---|---|---|
| W/O Iterations | | With Iterations | | |
| Bayesian | SVM | Iterations | BB | EGP |
| 85 | 86.4 | 1 | 85 | 80.4 |
| | | 5 | 85.1 | 82.3 |
| | | 10 | 85.1 | 85.7 |
| | | 15 | 85.1 | 85.1 |
| | | 20 | 85.1 | 85 |
| | | 25 | 85.1 | 86.5 |
| | | 30 | 85.1 | 86.2 |
| | | 35 | 85.1 | 86.3 |
| | | 40 | 85.1 | 86.4 |

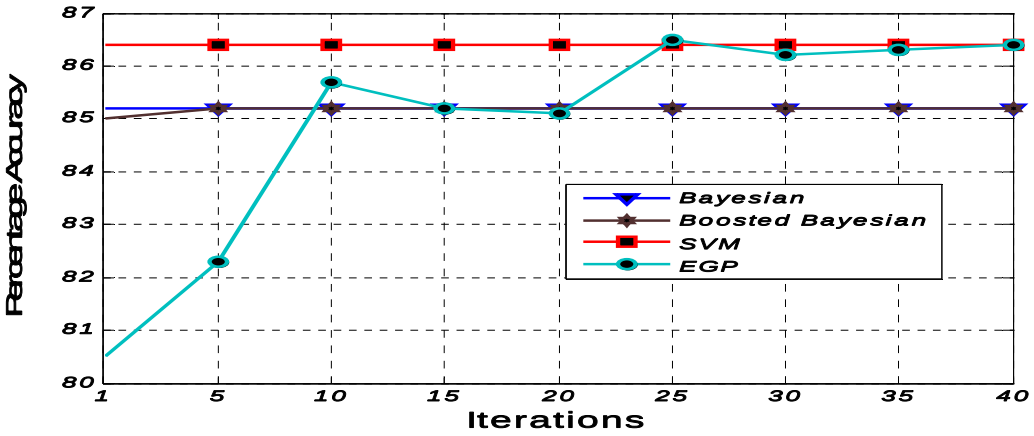


Fig 57. Percentage Accuracy for Machine Learning Classifiers

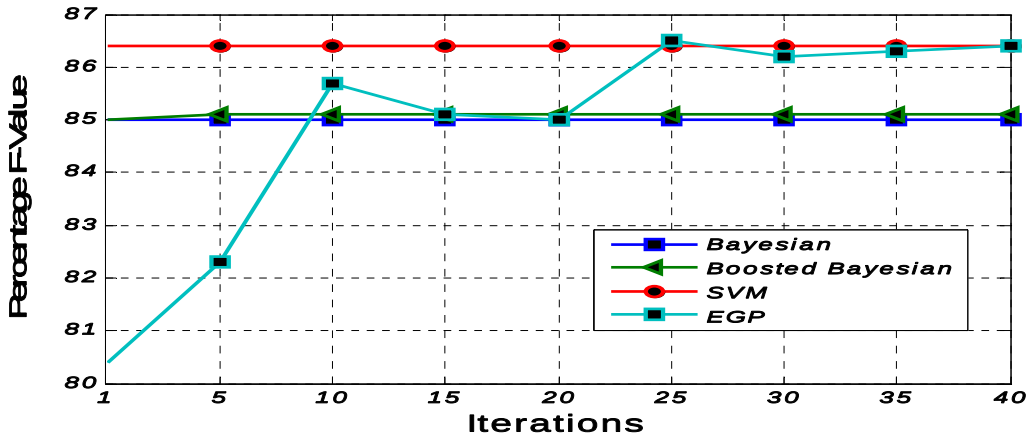


Fig 58. Percentage F-Value for Machine Learning Classifiers

Observation 1:

After analysing the results (Table 44, 45 and Figure 57, 58), the first observation shows that Support Vector Machine (*SVM)* is found to be the best classifier whereas Enhanced Genetic Programming (*EGP*) initially starts with low performance accuracy but after ensemble of 25 Weak Genetic Programming classifiers, the results overshoots to the best one. At the ensemble of 25 weak GP Classifiers, EGP is comparable to the SVM. In this time, the performance accuracy for EGP is 86.5% and for SVM is 86.4%.

Observation 2:

The second observation is for Bayesian and Boosted Bayesian (*BB*) Classifiers where both have given more or less similar results with 85.1% and 85.2% classification accuracy respectively (Table 44, 45 and Figure 57, 58). This time, small boosting has been observed from Boosted Bayesian (*BB*) classifier.

Observation 3:

In this observation, comparison between both ensembles of classifiers i.e. Boosted Bayesian (*BB*) and Enhanced Genetic Programming (*EGP*) has been analyzed. Results show that initially *EGP* classifier gives weak performance but after ensemble of 10 weak *GP* classifiers *EGP* overshoots the performance of *BB*. Classification accuracy of *BB* and *EGP* are 85.2% and 80.55%-86.5% respectively (Table 44, 45 and Figure 57, 58).

Observation 4:

Finally, when analysis is done with all classifiers together, then *EGP* and *SVM* is identified to be best. In addition, *EGP* classifier overshoots the performance of *SVM* when the ensemble is done with 25 *GP* classifiers.

*8.1.7.2 Analysis with False Positive Rate:*

The results of False Positive Rate are shown in Table 46 and Figure 59, from which some observations are identified.

Observation 1: The results show that Boosted Bayesian (*BB*) gives excellent sensitivity for accurate classification with low False Positive Rate (3.5%) after 5 iterations (number of ensemble of classifiers). In addition, Bayesian classifier is the second best with 3.8% False Positive Rate. In this time, *SVM* has been predicted to be worst with 6.1% FP Rate.

Observation 2: When the analysis is done between ensemble based classifiers i.e. *BB* and *EGP*, again *BB* shows to be excellent whereas the results of *EGP* are varying with the iterations.

Table 46. False Positive Rate for Machine Learning Classifiers

| Percentage FP rate | | | | |
|---|---|---|---|---|
| W/O Iterations | | With Iterations | | |
| Bayesian | SVM | Iterations | BB | EGP |
| 3.8 | 6.1 | 1 | 3.8 | 3.1 |
| | | 5 | 3.6 | 3.4 |
| | | 10 | 3.5 | 6.1 |
| | | 15 | 3.5 | 4.8 |
| | | 20 | 3.5 | 6.9 |
| | | 25 | 3.5 | 5.6 |
| | | 30 | 3.5 | 4.3 |
| | | 35 | 3.5 | 6 |
| | | 40 | 3.5 | 5.8 |

Observation 3:

The observation that is found from *EGP* and *SVM* show that for all iterations (with the ensemble of 40 weak classifiers) the False Positive Rate of *EGP* classifier is lower than *SVM.*

.

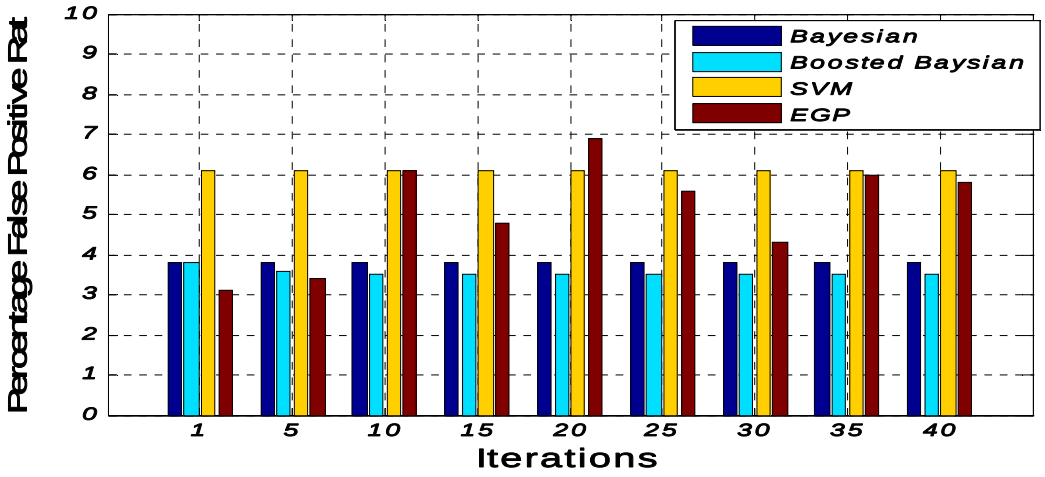


Fig 59. Percentage FP Rate for Machine Learning Classifiers

Table 47. Training Time for Machine Learning Classifiers

| Training Time (In Second) | | | | |
|---|---|---|---|---|
| W/O Iterations | | With Iterations | | |
| Bayesian | SVM | Iterations | BB | EGP |
| 0.3 | 1.92 | 1 | 0.39 | 28.79 |
| | | 5 | 3.64 | 132.75 |
| | | 10 | 3.64 | 275.9 |
| | | 15 | 6.49 | 410.9 |
| | | 20 | 6.68 | 547.13 |
| | | 25 | 9.15 | 687.13 |
| | | 30 | 9.35 | 839.96 |
| | | 35 | 10.74 | 982.47 |
| | | 40 | 12.4 | 1079.52 |

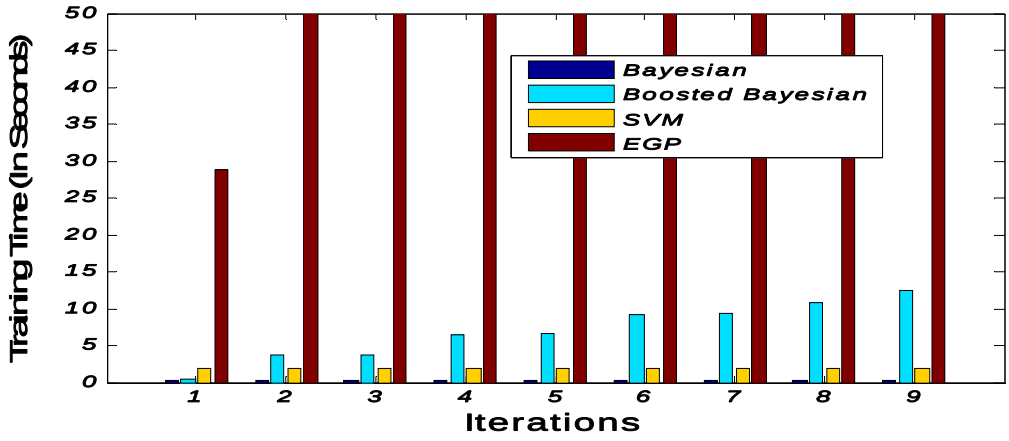


Fig 60. Training Time for Machine Learning Classifiers

*8.1.7.3 Analysis with Training Time:*

Training Time is now becoming a primary concern for organisations and Internet Service Providers (ISPs). A major trade-off between Classification Accuracy and Training Time has been identified in this study. It is observed that for training of ensemble of classifiers model, huge time is required. Another analysis suggests that once a filtering system is trained with predefined examples then only Testing Time is necessary. The results of Training Time of the concerned classifiers are shown in Table 47 and Figure 60 which identify three different observations. First observation compares Training Time of the classifiers without ensemble, second takes the analysis on ensemble based classifiers and last, all classifiers are taken together for comparing the training time.

Observation 1:

The results of Training Time for the classifiers (without ensembles) show that the training time for Bayesian Classifier is less i.e. 0.3 sec whereas *SVM* (Normalized Poly Kernel) takes more Training Time that is 1.92 sec.

Observation 2:

The observations for Ensemble based classifiers i.e. Boosted Bayesian (*BB*) and Enhanced Genetic Programming (*EGP*) suggest that Boosted Bayesian is trained rapid than *GP*. As soon as the weak classifiers are added for developing ensembles, the classifiers model take more time in training.

Observation 3:

Among all classifiers, Bayesian has been predicted to be quick with 0.3 sec Training Time whereas *SVM* comes in the second position with 1.92 sec. Boosted Bayesian (*BB*) takes third position with 0.39 sec (with one weak classifier) and 12.4 sec (for ensemble of 40 weak classifiers). Training time for EGP is huge and predicted to be the worst classifier in training.

*8.1.7.4 Analysis with Testing Time:*

Although, Training Time is necessary for developing a cost sensitive classification model but Testing Time is also important to be captured because for a robust and sensitive classifier, rapid testing is necessary for real time environment. The results of Testing Time are shown in Table 48 and Figure 61 from which some observations are developed:

Observation 1:

This observation is derived from comparing the results of the classifiers without ensembles. In this case, Bayesian classifiers are identified to be rapid with 2.0 sec Testing Time whereas *SVM* is predicted to be worst with 28.3 sec Testing Time.

Table 48. Testing Time for Machine Learning Classifiers

| Training Time (In Second) | | | | |
|---|---|---|---|---|
| W/O Iterations | | With Iterations | | |
| Bayesian | SVM | Iterations | BB | EGP |
| 2 | 28.3 | 1 | 1.8 | 0.9 |
| | | 5 | 2.2 | 1.8 |
| | | 10 | 3.6 | 4.5 |
| | | 15 | 4.7 | 6.2 |
| | | 20 | 6.1 | 8.6 |
| | | 25 | 7.8 | **10.6** |
| | | 30 | 9.2 | 13.5 |
| | | 35 | 10.4 | 15.2 |
| | | 40 | 11.9 | **16.3** |

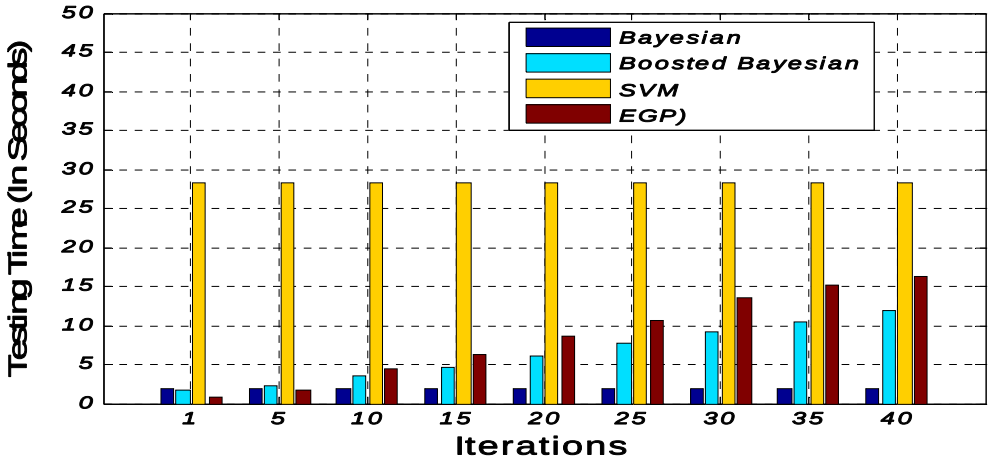


Fig 61. Testing Time for Machine Learning Classifiers

Observation 2:

When the analysis is done with ensemble of classifiers, initially *EGP* classifier for 1 to 5 iterations (number of weak classifiers taken for ensemble) gives fast testing with 0.9 sec to 1.8 sec Training Time. As long as more classifiers are included for ensemble, Testing Time increases from 1.8 sec to 16.3 sec. On the other side, *BB* classifier initially gives slower testing than *EGP* with 1.8 sec to 2.2 sec Testing Time but after 5 iterations *BB* do fast testing than *EGP*.

Observation 3:

When the analysis is done between *SVM* and *EGP (*identified as most accurate classifiers), *EGP* with all iterations dominates on *SVM* with low Testing Time.

*8.1.7.5 Analysis with all Metrics:*

In this section, when all the metrics are taken together for analyzing robust and rapid classifiers, two main observations have been formulated:

Observation 1:

This observation favors those organisations and Internet Service Providers whose primary concern is to reduce Training Time. For such organisations, Bayesian and Boosted Bayesian (*BB*) could be an excellent choice because it has less Training and Testing time but classification accuracy was less than other classifier compared.

Observation 2:

This observation benefits those organisations and Internet Service Providers (ISPs) whose main concern is to improve classification accuracy but they do not bother about testing time. For such scenario, *EGP* classifier would be an excellent choice because in $25^{th}$ iteration it has given best classification with low False Positives and rapid testing time but with suboptimal Training Time.

## 8.2 Discussion:

At the beginning of this study, we have intended to construct a robust, fast, sensitive and accurate email filtering system that could cater all the basic need of the organization and service providers. The aim of this study was successfully achieved by developing two different classification models. The first model has been developed by Enhanced Genetic Programming (EGP) classifier with Greedy Stepwise Feature Subset Selection method that was accurate, sensitive with low false positive, fast in training but testing time was more amongst the other models compared in this study. On the other hand, second model has been constructed by the Bayesian classifier with Greedy Stepwise Feature Subset Selection method that was sensitive with less false positive, fast training and testing but the accuracy was less than the first model.