CS571 AI LAB 08

13th November 2022

Ishita Singh 1901CS27
Kavya Goyal 1901CS30
Priyanka Sachan 1901CS43

Team Code: 1901cs30_1901cs27_1901cs43
Collab Link: cs30_cs27_cs43_AI08

Designing a spam filtering based on Naive Bayes classifier.

- Implement multinomial and multivariate
- Add One Smoothing

#5 fold cross validations

```
    if x == 0:
        correct_predictions_spam = (tp) / (tp + fn)
        correct_predictions_ham = (tn) / (tn + fp)
        correct_predictions = (tp + tn) / (tp + fp + tn + fn)
        return (correct_predictions_spam, correct_predictions_ham, correct_predictions, time_taken)
```

#10 fold cross validations

```
    else:
        precision = (tp) / (tp + fp)
        recall = (tp) / (tp + fn)
        return (precision, recall, time_taken)
```

Accuracies:

```
# calculating results in ter,s of category wise accuracies and overall accuracies
accuracy_spam_MN /= k
accuracy_ham_MN /= k
accuracy_MN /= k
accuracy_spam_MV /= k
accuracy_ham_MV /= k
accuracy_MV /= k
time_MN /= k
time_MV /= k
```

RESULTS FOR 5 FOLD CROSS VALIDATION :

Result	Value
Accuracy_SPAM_MN	0.8785414779439339
Accuracy_HAM_MN	0.9964741479599557
Accuracy_MN	0.9807899461400359
AvgTime_MN	0.5669358253479004
Accuracy_SPAM_MV	0.9838122620730223
Accuracy_HAM_MV	0.8150038412844476
Accuracy_MV	0.83770197486535
AvgTime_MV	0.5533849239349365

RESULTS FOR 10 FOLD CROSS VALIDATION :

Result	Value
Precision_MN	0.9766815870417277
Recall_MN	0.8935313426759317
F1_Score_MN	0.9332580220884295
Precision_MV	0.4714178530548582
Recall_MV	0.9865940998403037
F1_Score_MV	0.6379893819934263

Why multinomial is better than multivariate in terms of precision:

In multivariate, the classification does not take into account the effect of the frequency of words occurring in ham and spam, where as in multinomial, the effect of a word and its frequency of being in ham/spam class is also taken into account