# Document Categorization or Text Classification

**Dataset**

We have taken a dataset with 5000 paragraphs in Bangla language. They are of different sizes. Each of the 5000 paragraph is represented by 34,465 features.

## Results (accuracy)

1. LinearSVC - 72.63%

2. LogisticRegression– 67.87%

3. MultinomialNB – 42.23%

4. RandomForestClassifier – 26.08%

# Sentiment Analysis

**Dataset**

We have taken a dataset with 27731 items in Bangla language. Each of the 27731 paragraph is represented by 20759 features.

## Results (accuracy)

1. LinearSVC - 23.90%
2. LogisticRegression – 39.33%
3. MultinomialNB – 39.33%
4. RandomForestClassifier – 40.00%