2020503

```
┌─────────────────────────┐          ┌─────────────────────────┐
│ 20 Newsgroup datasets   │          │  Reuters   datasets     │
└─────────────────────────┘          └─────────────────────────┘
              │                                  │
              └────────────────┬─────────────────┘
                               ▼
              ┌───────────────────────────────┐
              │ Extracting features and       │
              │ labels for each training      │
              │ and testing datasets          │
              └───────────────────────────────┘
                               │
                               ▼
                    ┌────────────────────┐
                    │ preprocessing step │
                    └────────────────────┘
                               │
                               ▼
                 ┌─────────────────────────┐
                 │ Splitting into training │
                 │ and testing datasets    │
                 └─────────────────────────┘
                               │
                               ▼
                 ┌─────────────────────────┐
                 │ Term Frequency and      │
                 │ Inverse document        │
                 │ Frequency               │
                 └─────────────────────────┘
                               │
                               ▼
                 ┌─────────────────────────┐
                 │ Model Implementation    │
                 │ from sklearn            │
                 └─────────────────────────┘
                               │
        Accury                 ▼
                 ┌─────────────────────────┐
                 │ Accuracy, precision     │
                 │ Recall, F1 Score        │
                 └─────────────────────────┘
                               │
                               ▼
                 ┌─────────────────────────┐
                 │ Finding which model     │
                 │ is good on behalf of    │
                 │ Accuracy precision      │
                 │ recall and F1 Score     │
                 └─────────────────────────┘
                               │
                               ▼
                 ┌─────────────────────────┐
                 │ visualization of        │
                 │ all accuracies          │
                 │ precision, Recall       │
                 │ and F1 Score            │
                 └─────────────────────────┘
```
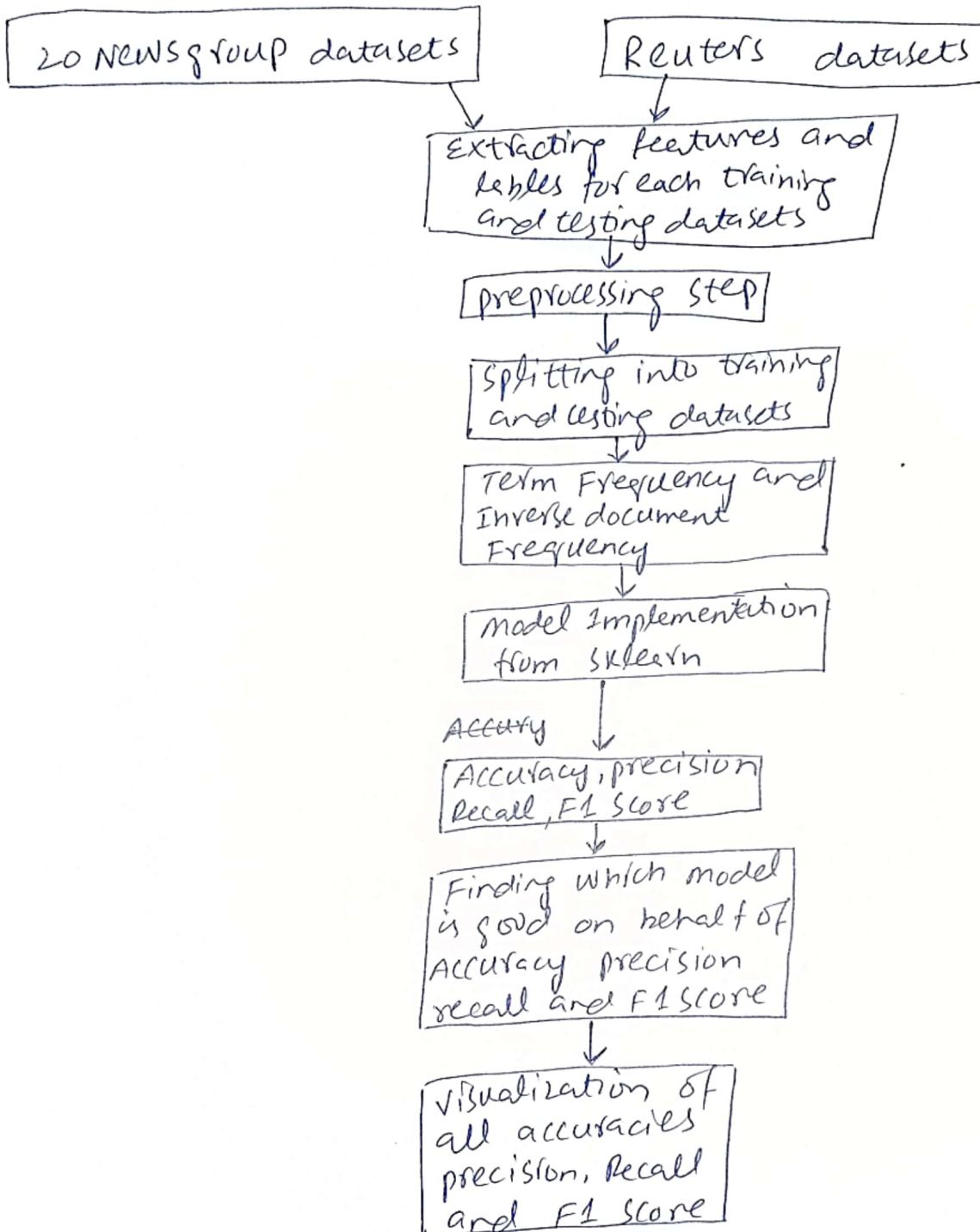
Based on our Experimental Analysis we have found that logistic regression outperformed other models for text classification on the 20 News groups and Reuters News Datasets.
Now its coming to reason why logistic Regression performed better then other model's.

1) Linear separability
2) Less prone to overfitting
3) Better Handling of Imbalanced data
4) Efficient Training