

Data Science - Shared Language Identification Task

RESULTS AND CONCLUSIONS

The attached folder contains:

3 - python files:

1. final.py - user interactive code which predicts tags for sentence given as input by user
2. task1.py - Week1 Task
3. task2.py - Week2 Task i.e., with CRF

Running Instructions:

```
>>>python final.py
```

```
>>>python task1.py
```

```
>>>python task2.py
```

Report on F1_Scores:

Strict F-measure of classifiers versus languages

Language	NaiveBayes	MaxEntropy
English	0.82	0.84
Bengali	0.85	0.84
Gujarathi	0	0
Hindi	0.73	0.78
Kannada	0.77	0.77
Malayalam	0.80	0.79
Marati	0.88	0.91
Tamil	0.80	0.79
Telugu	0.59	0.26

Strict F-measure of classifiers versus languages with CRF

Language	NaiveBayes	MaxEntropy
English	0.86	0.85
Bengali	0.85	0.85
Gujarathi	0	0
Hindi	0.75	0.77
Kannada	0.76	0.77
Malayalam	0.81	0.79
Marati	0.88	0.91
Tamil	0.79	0.80
Telugu	0.60	0.53

Average F-Scores of Classifiers:

Method	Average F-Score
MaxEntropy	0.7761
Naive Bayes	0.7860
MaxEntropy+CRF	0.7973
Naive Bayes+CRF	0.7930

*All the python notebooks in the folder are practise based

*The codes final.py, task1.py, task2.py works only in the submitted folder directory as path was not mentioned.

Submitted By:-

Group as per pair programming

1) Sri Harshini.V (411780)

2) Sai.A (411783)