

# Indian Language Identification Model

## Report

- Malhar Brahmbhatt

Total examples:

79,283

Sentence counts in training data:

Kashmiri\_Arab: 100000

Kannada: 100000

Nepali: 115554

Manipuri\_Mei: 100000

Hindi: 100000

Tamil: 100000

Telugu: 100000

Oriya: 100000

Urdu: 105282

Manipuri\_Beng: 100000

Bodo: 108240

Marathi: 100000

Sanskrit: 102315

Assamese: 100000

Maithili: 100277

Punjabi: 100000

Gujarati: 100000

Other: 100000

Bangla: 100000

Malayalam: 100000

Dogri: 100000

Konkani: 100000

Kashmiri\_Deva: 100000

Sindhi: 100000

English: 100000

Santali: 100000

### Sentence counts in validation data:

Kashmiri\_Arab: 1497  
Bangla: 5997  
Tamil: 5997  
Manipuri\_Beng: 997  
Marathi: 5997  
Sindhi: 5996  
Gujarati: 5997  
Malayalam: 5997  
Punjabi: 5997  
Telugu: 5997  
Hindi: 5997  
Assamese: 997  
Kannada: 5997  
Urdu: 6376  
Oriya: 997  
Maithili: 1497  
Sanskrit: 1497  
English: 997  
Manipuri\_Mei: 500  
Nepali: 1497  
Konkani: 500  
Bodo: 500  
Dogri: 120  
Kashmiri\_Deva: 997  
Santali: 345

### Accuracy:

Overall accuracy:

0.9315

Language-wise accuracy:

Accuracy for Kashmiri\_Arab: 0.9866

Accuracy for Bangla: 0.9473  
Accuracy for Tamil: 0.9218  
Accuracy for Manipuri\_Beng: 0.9829  
Accuracy for Marathi: 0.8549  
Accuracy for Sindhi: 0.9823  
Accuracy for Gujarati: 0.9631  
Accuracy for Malayalam: 0.9256  
Accuracy for Punjabi: 0.9747  
Accuracy for Telugu: 0.8891  
Accuracy for Hindi: 0.8573  
Accuracy for Assamese: 0.9970  
Accuracy for Kannada: 0.9538  
Accuracy for Urdu: 0.8781  
Accuracy for Oriya: 0.9990  
Accuracy for Maithili: 0.9345  
Accuracy for Sanskrit: 0.9719  
Accuracy for English: 0.9930  
Accuracy for Manipuri\_Mei: 1.0000  
Accuracy for Nepali: 0.9893  
Accuracy for Konkani: 0.9880  
Accuracy for Bodo: 0.9860  
Accuracy for Dogri: 0.9917  
Accuracy for Kashmiri\_Deva: 0.9408  
Accuracy for Santali: 1.0000