Language Recognition

```
In [13]: %store -r accuracy_df
         accuracy_df
```

Out[13]:

		Training Data Accuracy	Test Data Accuracy
Count Vectorizer	Logistic Regression	0.993295	0.939773
	Naive Bayes	0.991023	0.959318
Tf-idf	Logistic Regression	0.974034	0.941364
	Naive Bayes	0.983920	0.939545

```
In [3]: %store -r cross_val_scores
        cross_val_scores
```

Out[3]:

		Cross Validation Score
Count Vectorizer	Logistic Regression	0.947773
	Naive Bayes	0.955409
Tf-idf	Logistic Regression	0.955000
	Naive Bayes	0.954545

First Sentence Prediction

```
In [14]: %store -r accuracy_f1
         accuracy_f1
```

Out[14]:

		Training Data	Test Data
Count Vectorizer	Accuracy	0.673437	0.533291
	F1 Score	0.645661	0.443936
Tf-idf	Accuracy	0.695176	0.592190
	F1 Score	0.684098	0.547264

```
In [4]: %store -r tf_svd_25_50_75_100
        tf_svd_25_50_75_100
```

Out[4]:

		Training Data	Test Data
25 components	Accuracy	0.665799	0.618821
	F1 Score	0.655820	0.630415
50 components	Accuracy	0.662326	0.629753
	F1 Score	0.645138	0.638850
75 components	Accuracy	0.689931	0.621673
	F1 Score	0.678198	0.619139
100 components	Accuracy	0.659549	0.614068
	F1 Score	0.628669	0.565310

```
In [2]:  %store -r tf_optimized
         tf_optimized
```

Out[2]:

		Training Data	Test Data
Un-optimized	Accuracy	0.662326	0.629753
	F1 Score	0.645138	0.638850
Optimized for F1	Accuracy	0.662326	0.629753
	F1 Score	0.645138	0.638850

```
In [2]:  %store -r tf_optimized
         tf_optimized
```

Out[2]:

		Training Data	Test Data
Un-optimized	Accuracy	0.662326	0.629753