Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis

Replacing Negative mentions:

Gain/Lost	Feature Model	Classifier	STS-Test	
			Binary	3 way
Accuracy	Prior Polarity	-	-	0.0
	N-grams	Logistic Regression	0.06	-
		Naive Bayes	0.03	0.03
		SVM	0.05	0.06
		Random Forest	0.0	-0.11
F1 - Score	Prior Polarity	-	-	-0.01
	N-grams	Logistic Regression	0.08	-
		Naive Bayes	0.05	0.04
		SVM	0.05	0.03
		Random Forest	0	-0.1

Removing URLs:

Gain/Loss	Feature Model	Classifier	STS-Test	
			Binary	3 way
Accuracy	Prior Polarity	-	-	0.0
	N-grams	Logistic Regression	0.01	-
		Naive Bayes	0.03	0.03
		SVM	-0.02	0.04
		Random Forest	-0.05	-0.1
F1 - Score	Prior Polarity	-	-	0.0
	N-grams	Logistic Regression	0.0	-
		Naive Bayes	0.01	0.05
		SVM	-0.04	0.06
		Random Forest	-0.03	0.01

Reverting repeating characters:

Gain/Loss	Feature Model	Classifier	STS-Test	
			Binary	3 way
Accuracy	Prior Polarity	-	-	-0.04
	N-grams	Logistic Regression	0.0	-
		Naive Bayes	0.06	0.1
		SVM	-0.01	0.12
		Random Forest	-0.07	-0.07
F1 - Score	Prior Polarity	-	-	-0.07
	N-grams	Logistic Regression	0.05	-
		Naive Bayes	0.03	0.16
		SVM	-0.05	0.15
		Random Forest	-0.05	-0.01

Removing stopwords:

Gain/Loss	Feature Model	Classifier	STS-Test	
			Binary	3 way
Accuracy	Prior Polarity	-	-	0.0
	N-grams	Logistic Regression	0.0	-
		Naive Bayes	0.0	0.01
		SVM	-0.03	0.04
		Random Forest	-0.03	-0.09
F1 - Score	Prior Polarity	-	-	-0.01
	N-grams	Logistic Regression	0.0	-
		Naive Bayes	-0.01	0.04
		SVM	-0.03	0.04
		Random Forest	-0.06	-0.07

Replacing acronyms:

Gain/Loss	**Feature Model**	**Classifier**	**STS-Test**	
			Binary	**3 way**
Accuracy	Prior Polarity	-	-	0.0
	N-grams	Logistic Regression	-0.02	-
		Naive Bayes	0.02	0.04
		SVM	-0.01	0.0
		Random Forest	-0.06	-0.15
F1 - Score	Prior Polarity	-	-	-0.01
	N-grams	Logistic Regression	-0.01	-
		Naive Bayes	0.02	0.09
		SVM	-0.02	0.02
		Random Forest	-0.05	-0.08