



# Tweet Classification

MURALI KRISHNA VALLURI (MVALLU2@UIC.EDU)

SPOORTHY PENDYALA (NPENDY2@UIC.EDU)

# Pre - Processing Steps

- ▶ Initial excel sheet is converted to csv file by removing rows with class values '2', 'IR', 'Irrelevant', '!!!!'. Also Date and Time columns were removed.
- ▶ Tweet is converted to lowercase.
- ▶ Removed special characters and replaced '#', '@', '"', ''' with blank characters.
- ▶ Removed hyperlinks and html tags.
- ▶ Stripped the text which removes extra spaces at the beginning and end of the text.
- ▶ Finally we removed the stop words from the text.

# Classification Algorithms

- ▶ Decision Trees
- ▶ Logistic Regression
- ▶ Multinomial Naïve Bayes
- ▶ K – Nearest
- ▶ SVM (No results were obtained for positive class, hence we did not consider this algorithm for the final comparison.)

**Note:** The results were obtained via 10-fold cross validation on the training data.

# Results

## ► Obama Data

Algorithms	Negative Class			Positive Class			Accuracy
	Precision	Recall	F-Score	Precision	Recall	F-Score	
Logistic Regression	0.57	0.68	0.61	0.64	0.58	0.61	58.47
Multinomial NB	0.53	0.79	0.62	0.67	0.54	0.59	56.79
K-Nearest	0.66	0.29	0.37	0.59	0.50	0.52	47.37
Decision Tree	0.49	0.45	0.46	0.45	0.49	0.47	45.81

## ► Romney Data

Algorithms	Negative Class			Positive Class			Accuracy
	Precision	Recall	F-Score	Precision	Recall	F-Score	
Logistic Regression	0.59	0.88	0.69	0.64	0.28	0.39	57.21
Multinomial NB	0.54	0.99	0.68	0.76	0.9	0.16	53.68
K-Nearest	0.53	0.95	0.67	0.71	0.5	0.10	52.54
Decision Tree	0.58	0.61	0.58	0.36	0.32	0.34	47.97