Twitter Sentimental Analysis

CS 583 Research Project – Spring 2017

Keshav Malpani – kmalpa2

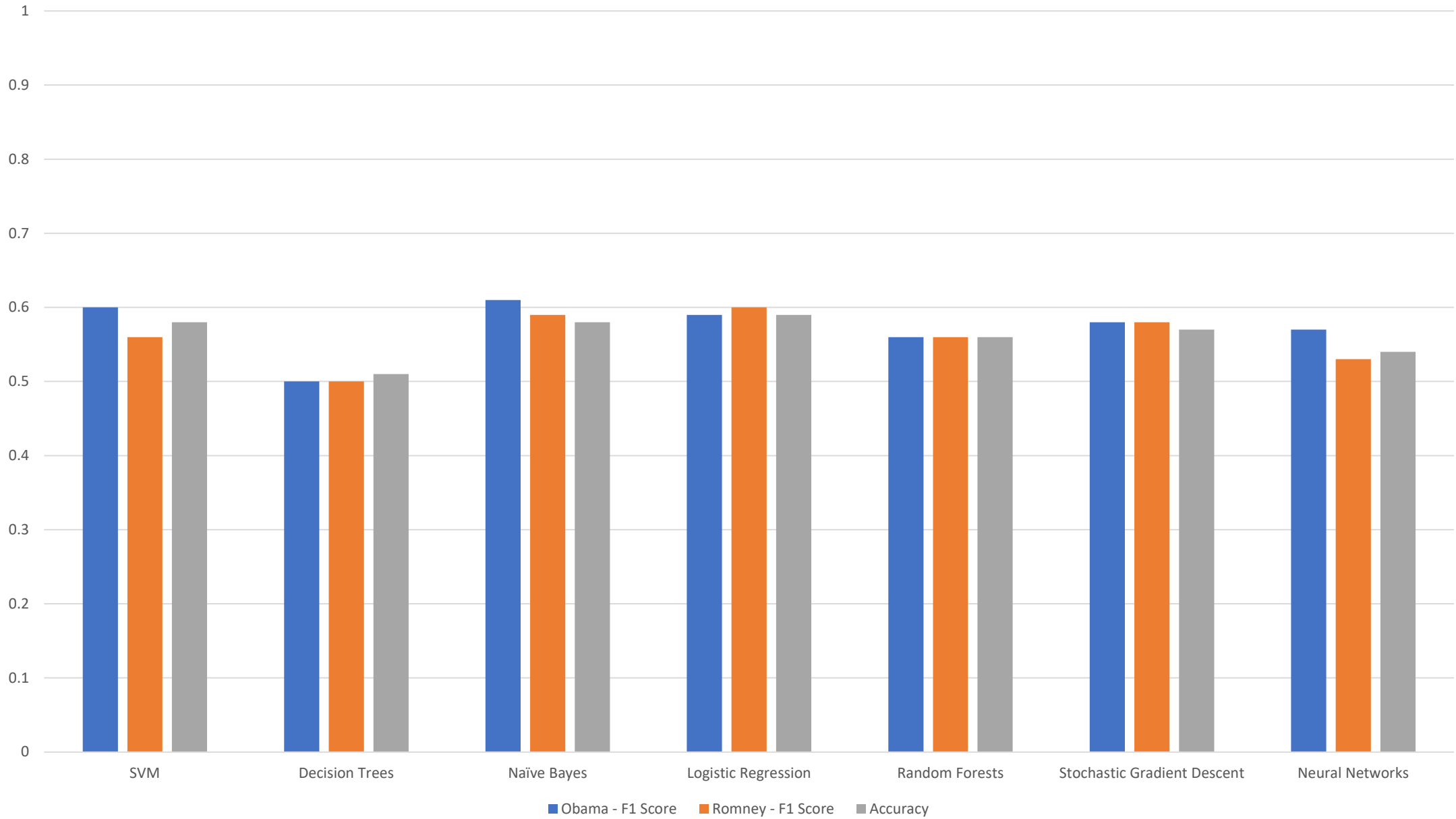
Venkata Ramkiran Chevendra – vcheve2

Data Preprocessing

- Data cleaning
  - Removed usernames, URL, HTML tags, hashtags, punctuation, stop words and special characters
  - Stemming
  - Slang Conversion
  - Tokenization
  - Tf-idf Vectorization
- Training and Testing
  - K – Fold Cross Validation (k = 10)
  - Basic data split (90% Train & 10% Test)

Classifiers and Results

Model	Overall Accuracy	Candidate	Positive Class			Negative Class			Average F1 - Score	Accuracy
			Precision	Recall	F1 - Score	Precision	Recall	F1 - Score		
SVM	0.58	*Obama*	0.6	0.55	0.58	0.6	0.63	0.62	0.6	0.57
		Romney	0.74	0.28	0.41	0.58	0.93	0.71	0.56	0.59
Decision Trees	0.51	*Obama*	0.45	0.51	0.47	0.57	0.49	0.52	0.5	0.5
		Romney	0.4	0.33	0.36	0.6	0.67	0.63	0.5	0.51
Naïve Bayes	0.58	*Obama*	0.56	0.61	0.58	0.66	0.59	0.62	0.61	0.57
		Romney	0.56	0.46	0.5	0.62	0.76	0.68	0.59	0.58
Logistic Regression	0.59	*Obama*	0.59	0.55	0.57	0.62	0.6	0.6	0.59	0.56
		Romney	0.68	0.35	0.46	0.62	0.88	0.73	0.6	0.61
Random Forests	0.56	*Obama*	0.53	0.51	0.52	0.6	0.56	0.58	0.56	0.53
		Romney	0.54	0.34	0.42	0.63	0.81	0.71	0.56	0.58
Stochastic Gradient Descent	0.57	*Obama*	0.56	0.57	0.57	0.61	0.57	0.59	0.58	0.55
		Romney	0.6	0.38	0.47	0.63	0.79	0.7	0.58	0.59
Neural Networks	0.54	*Obama*	0.54	0.58	0.56	0.6	0.57	0.58	0.57	0.54
		Romney	0.43	0.39	0.42	0.65	0.61	0.63	0.53	0.53


1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
SVM
Decision Trees
Naïve Bayes
Logistic Regression
Random Forests
Stochastic Gradient Descent
Neural Networks
Obama - F1 Score
Romney - F1 Score
Accuracy