

deletweet-introduction

April 3, 2017

1 DELETWEET

1.0.1 APPLYING TEXT MINING AND A RECOMMENDATION SYSTEM TO POLITICIAN'S DELETED TWEETS

1.1 INTRODUCTION

In the wake of the recent US presidential election, the humble tweet has been thrust to the forefront of social and political consciousness. 38.5 million people follow Donald Trump on Twitter via his two handles, @POTUS (the official acting US president's account) and @realDonaldTrump, and probably a billion more keep track of his tweets via the news media and other outlets. His casual remarks elicit palpable responses from financial markets, and influential public policy takes shape from seemingly off-the-cuff statements. In other words, these snippets of 140 characters or less have the power to move huge sums of money and influence millions of people's lives in real and profound ways.

In this climate, a weather report of sorts is called for. This project aims to examine tweets by politicians more closely, specifically those tweets that have been deleted. At our request [ProPublica](#) has generously provided their [Politwoops](#) dataset, which consists of tweets deleted by politicians and public officials in the United States. We aim to discern patterns in the text via machine learning, text mining, and natural language processing. We will also attempt to demonstrate a recommendation system that suggests relevant hashtags for a tweet.

1.2 QUESTIONS

This project applies text mining, natural language processing, and machine learning to answer 4 questions:

- * What are the most common topics in the dataset?
- * What are the most common sentiments or emotions expressed in the dataset?
- * Are the deleted tweets more likely to be authored at a particular time of day or day of the week?
- * Is it possible to recommend hashtags for a tweet based on its content?