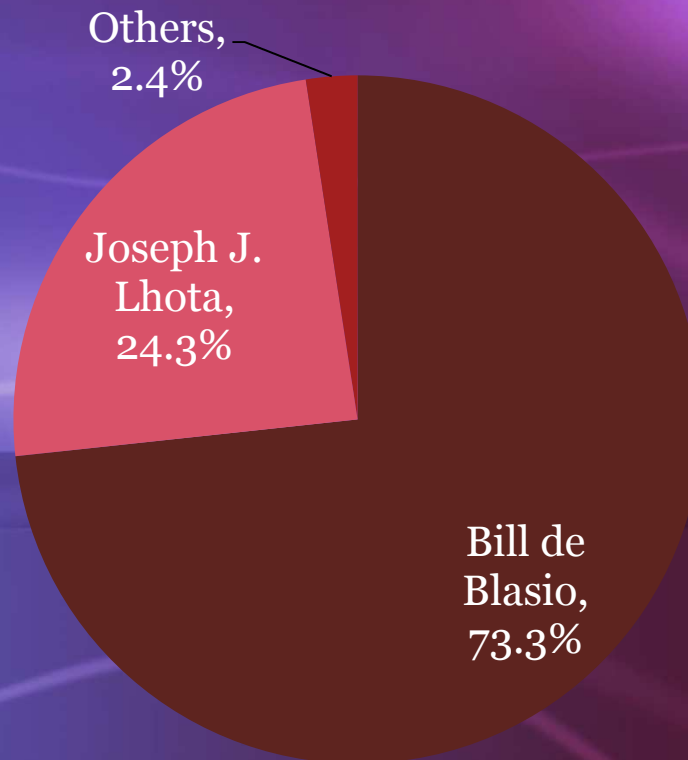


Using Twitter Sentiment to Predict the NYC Mayoral Election

November 14, 2013

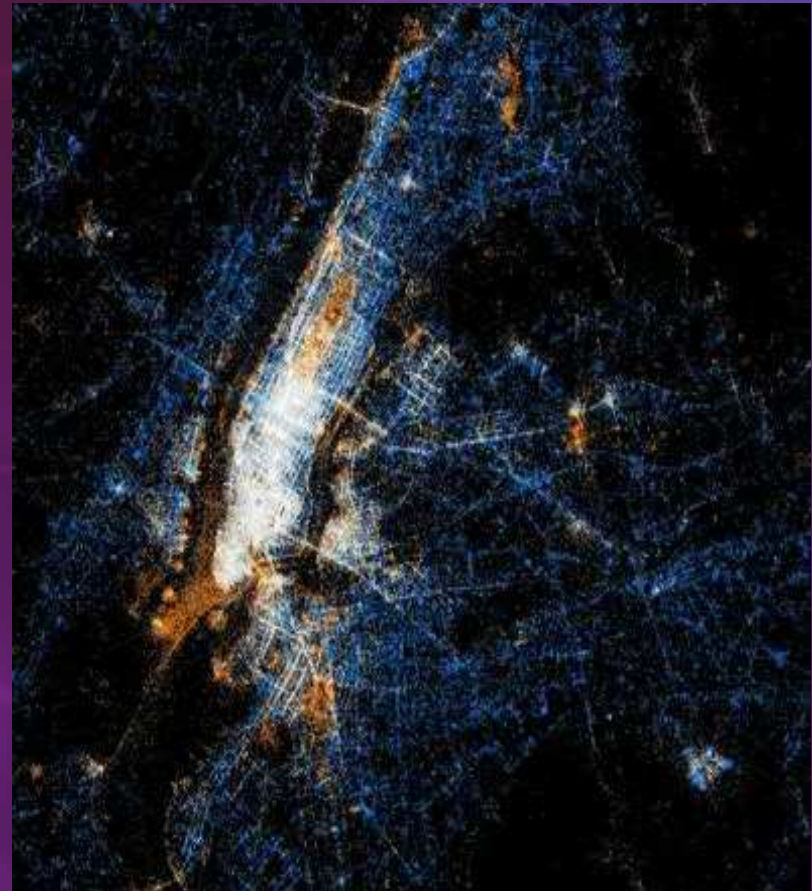
Election Outcome

Candidate	Votes	Percentage
Bill de Blasio (D)	752,604	73.3%
Joseph J. Lhota (R)	249,121	24.3%
Others	24,443	2.4%



Use Case

- “New York City has more Twitter users than any other city in the world and the second most Twitter developers” – Jack Dorsey
- New Yorkers are extremely active on Twitter: in 2009, while the region’s users owned only 1.44% of accounts, they created 2.37% of all tweets; they are collaborating in more complex ways than in other locations.



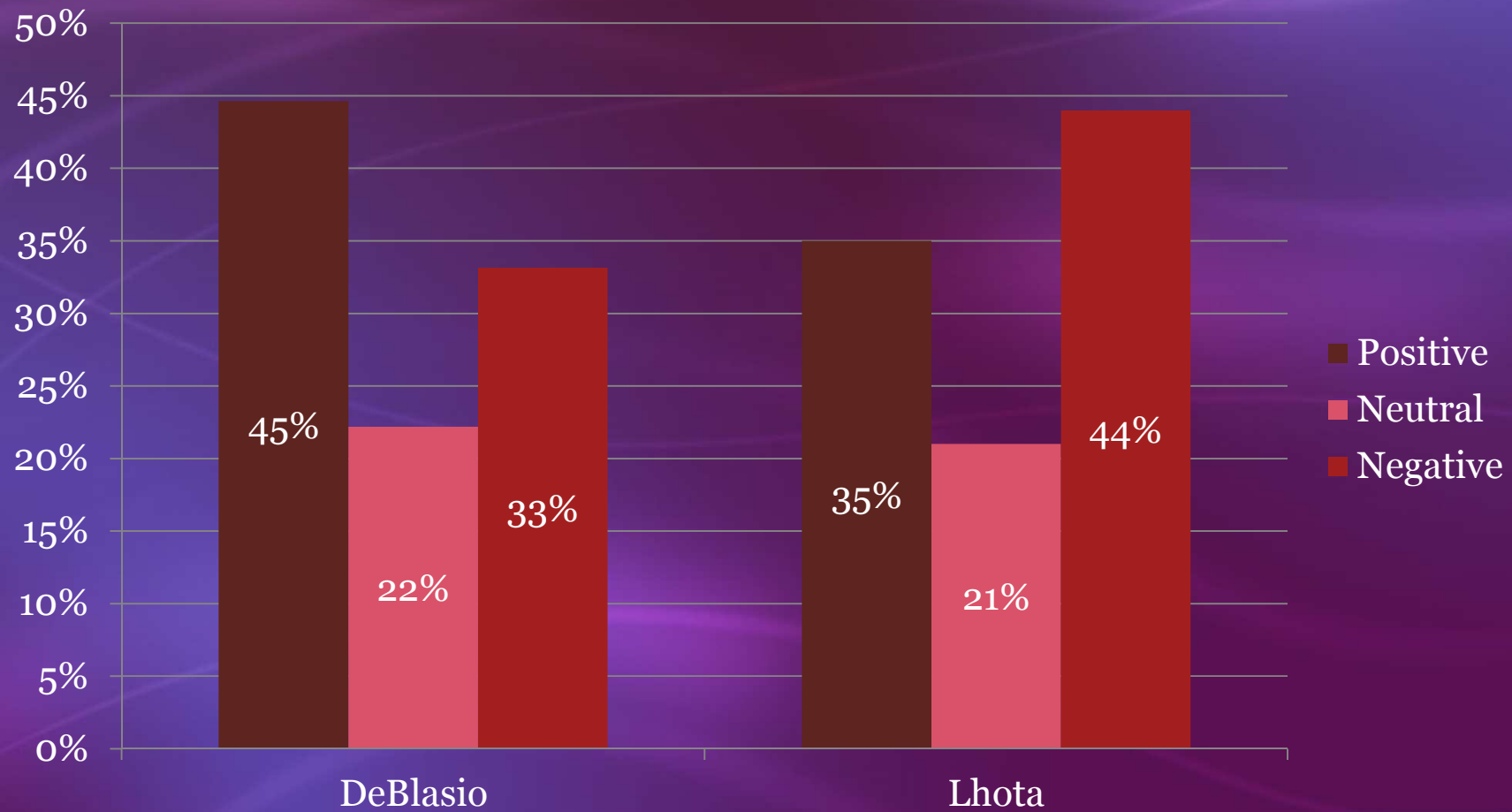
Process

- Obtained tweets from the last five days leading up to the election where 'Deblasio' or 'Lhota' was mentioned
- Classified tweet sentiment as positive, negative, or neutral
- Used Naïve Bayes to predict tweet sentiment
- Created topic models to understand what was being discussed most
- Analyze word frequency

Classifying Tweets

- Positive tweet:
 - SO EXCITED ABOUT BILL DEBLASIO
- Negative tweet:
 - Thanks Chicks On The Right!\nPoor Bill. He had to be up for an event at 11:30 am. That is, like, so harsh, dude
- Neutral tweet:
 - RT @ritchied: Could de Blasio vs Lhota be the most boring NYC mayoral race ever? Even their names sound like a collection of apathetic sigh

Distribution of Tweets



Using Naïve Bayes

- Not the greatest predictor
- Around 6,000 total tweets, about 5% were retweets
- Sarcasm or ambiguous tweet:
 - RT @AlexisinVT: Deblasio says he'll bring fundamental progressive change to #NYC Sound familiar? Wake up NYers! #nycmayor #tcot #teaparty #\u2026
 - #deblasio is not his real name. It's his mother's maiden name. Wilhelm Hitler is #deblasio 's real name. He is a descendant. @BilldeBlasio

Analysts Opinion on Most Important Issues

- Crime, the NYPD, and Stop and Frisk
- Labor Unions
- Education
- The Income Gap

Topic Modeling

- Top topics discussed for de Blasio:
 1. Stop
 2. Frisk
 3. Socialism
- Top topics discussed for Lhota:
 1. Giuliani
 2. Frisk
 3. Stop

Tweet Reactions

- On October 31, 2013 when the court blocked the order requiring changes to the NYPD stop and frisk program, 83% of all tweets that mentioned 'Debelasio' were about stop and frisk
- On October 31, 2013 when Rudy Giuliani stated he would campaign with Lhota, 43% of tweets were about Giuliani

For Lhota this was about 7% of tweets and for de Blasio this was about 10% of tweets.

Impact

- Can we use Twitter to determine outcomes?
- Can we use Twitter as a proxy to gauge reactions?
- Can we use Twitter to mitigate reactions?
 - Comparison of when a candidate responds to negative publicity vs. not responds
- Natural language processing techniques

Next Steps

- Analyze candidates reactions to tweets and overall activity
- Twitter social networks
- Classify tweets not in English
- Demographics
- Look at elections with multiple candidates and/or closer outcomes