AI Emotion Detector
and Topic Modeling in Suspended Elections Related Twitter Accounts.

happines
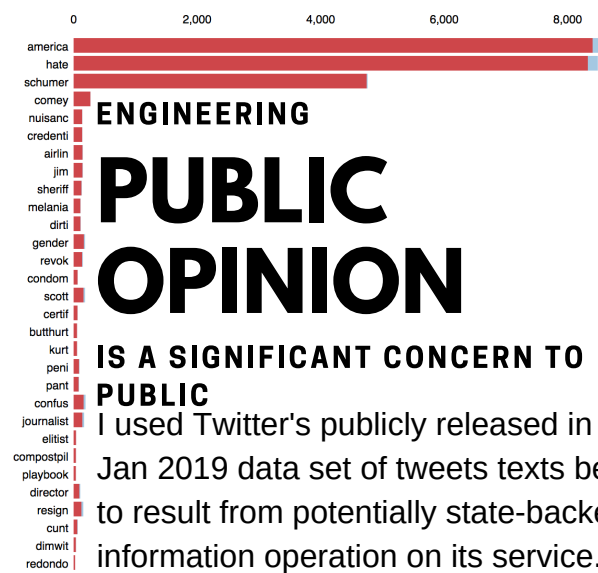fear
surprise
anger
sad
disgust

NASTYASIMPSON

Intertopic Distance Map (via multidimensional scaling)

PC2

PC1

3
6
7
5
4
2
1

Marginal topic distribution

2%
5%
10%

Top-30 Most Relevant Terms for Topic 3 (15.6% of tokens)

america
hate
schumer
comey
nuisanc
credenti
airlin
jim
sheriff
melania
dirti
gender
revok
condom
scott
certif
butthurt
kurt
peni
pant
confus
journalist
elitist
compostpil
playbook
director
resign
cunt
dimwit
redondo

ENGINEERING

PUBLIC OPINION

IS A SIGNIFICANT CONCERN TO PUBLIC

I used Twitter's publicly released in Jan 2019 data set of tweets texts believed to result from potentially state-backed information operation on its service.

Overall term frequency
Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
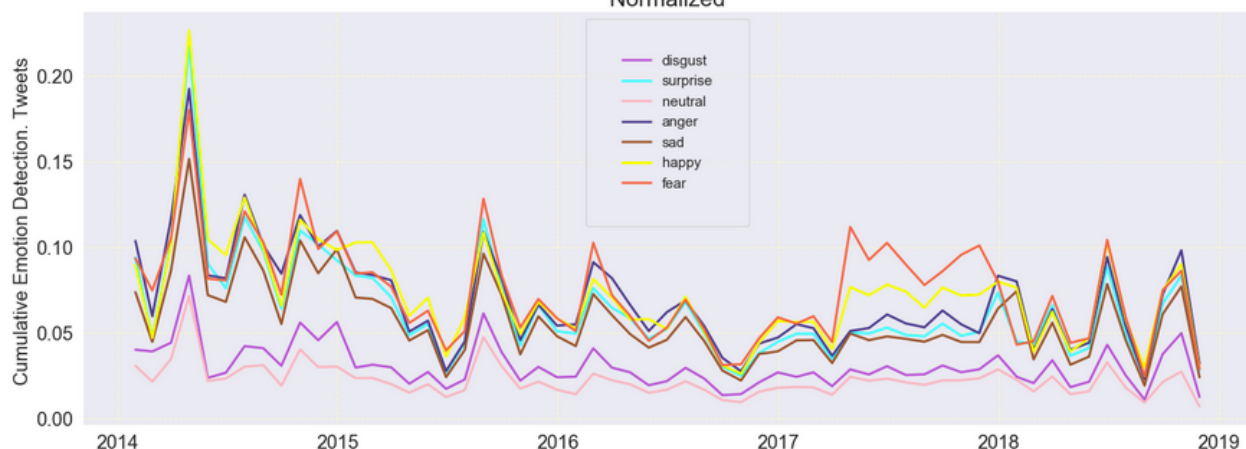2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Information retrieval is my passion.
This project was an investigation for hidden and latent features that resulted in building AI Emotion Detector using data set containing over 23,000 words statistically classified into 7 basic emotions.
After fitting my text data into the pipeline I built I observed that fear was predominant emotion of the data set with realtive increased fear before 2018 elections. Additionally I built Latent Dirichlet Allocation statistical model and observed that over 90% of text was distributed across American political issues. Race, Terror, Aggression, Disgust, Ashley, Madison, America, Hate, Shumer were among the top words.

pandas
NLTK NATURAL LANGUAGE PROCESSING
matplotlib
scikit learn
python
gensim topic modelling for humans

Detecting Seven Basic Emotions: 2014 -- 2019
Normalized

Cumulative Emotion Detection. Tweets

disgust
surprise
neutral
anger
sad
happy
fear

0.20
0.15
0.10
0.05
0.00

2014   2015   2016   2017   2018   2019

ANASTASIA SIMPSON
DATA SCIENTIST

anastasia-simpson   **nastyasimpson**   anastasiasimp@gmail.com   415-823-6731

SKILLS

Technical: Python, Pandas, Scikit-Learn, Matplotlib, Seaborn, HTML, SQL, G- Suite, Asana, Slack, git.

Statistics, Modeling & Machine Learning: Power Analysis, Bayesian Statistics, Decision Trees, Random Forest, AB Testing, Linear & Logistic Regression, kNN, Gradient Descent, ROC & Profit Curve, SVM, Boosting, NLP, NLTK, Naive Bayes, Gensim etc.

CASE STUDIES

Regression Case Study: Tractors prices prediction for enterprise. Data EDA and Feature Engineering. Used Linear Regression to model prices predictions.

Churn Prediction: Gradient Boosting Classifier. Archived 77 % accuracy on unseen test data. Proposed 4 different promotional ideas based on output of profit matrixes and ROC Curve.

Recommender System: Movies Recommender. Build a Logistic Regression model w/82 % accuracy.

Fraud Detection: Used NMF and K-Means, nlp, and tree models to predict Fraudulent Event Inquiries. Flask web app.

CAPSTONE PROJECT

AI Emotion Detector and Topic Modeling in Suspended Elections-Related Twitter Accounts. Gensim LDA, pyLDAvis, nlp, NLTK, geopandas, Twitter API, Twitter Sentiment and Emotion Analysis. Archived Topic Coherence Score of .7

EXPERIENCE   JAN 2019 — APR 2019

GALVANIZE, SAN FRANCISCO, Data Science Immersive, **galvanize.com**

- 3-month, 700+ hour data science immersive training program covering core data science areas and skills including probability, statistics,, machine learning algorithms, and data visualization.
- Studies supervised and unsupervised machine learning, natural language processing (NLP), predictive modeling, statistical methods, A/B testing, gradient descent, decision trees, web scraping, boosting, clustering, profit curves, imbalanced classes, neural networks.

NOVELA, SAN FRANCISCO, ASSISTANT GENERAL MANAGER   **MAR 2017 — DEC 2017**

- Managed highly successful event venue and lounge in FiDi with 20+ employees. Presented company daily through multiple interactions with customers, vendors, investors etc. Facilitated and managed high profile corporate events during major SF business conferences.

DOSA, SAN FRANCISCO   **JUL 2010 — APR 2017**
MULTI UNIT ASSISTANT GENERAL MANAGER   JAN 2015 — APR 2017
MULTI UNIT MANAGER + MARKETING MANAGER + MAITRE D, HOST   JUL 2010 — JAN 2015

- Managed a team of 50+ including 5+ managers across two locations in San Francisco. Analyzed P&L and budgets weekly and proactively implemented solutions. Minimized excessive labor. Payroll.
- Implemented new technologies and partnerships for DOSA: Moved DOSA from OpenTable to SeatMe (Yelp) and significantly reduced associated costs. Grew on-demand online ordering business via partnering with Caviar (Square), Doordash, Eat24, Uber Eats. Increased effectiveness by incorporating Asana and Slack to the workplace as well as transitioned to Breadcrumb from Micros POS.
- Acted as an owner to respond to Yelp Reviews. Consistently kept 4.25 avg w/over 2000 reviews to date.
- Provided all necessary support for ownership to run a successful profitable business in San Francisco. Proved to be resourceful, solution-oriented, having a clear head on business objectives.

EDUCATION

- National Research Nuclear University MEPhI (Moscow Engineering Physics Institute)   **2007 — 2009**
  Masters Degree in Psychology (USCES Evaluated)
- Lomonosov Moscow State University , Psychology, Social Psychology   **2002 — 2007**