

# Datamining and Sentiment Analysis Using Twitter Data

Shengye Chen -200354388

Elizabeth Rayner -200365470

Alexander Biezenski - 200315435

# Basic Premise

- Twitter data to get the view of the general public
- For getting the sentiment relating to policies, politicians and marketing
- Twitter data is collected by a keyword to get sentiment related to that keyword that can be sorted by region

# Datamining

- Datamining is discovering patterns in data
- Using natural language processing to get a sentiment analysis that extracts data from text
- Anomaly detection and text mining

# Basics of Harvesting Twitter Data

- Uses package Tweepy
- Must become authorized by Twitter
- Streaming API versus Search API
- Filter endpoint versus sample endpoint
- Gathers real time tweets based on a list of filtered keywords
- Tweets must be flattened in order to be properly analyzed

# One tweet

```
{
  "created_at": "Wed Feb 08 22:00:05 +0000 2018",
  "id": 9803801722248786,
  "id_str": "9803801722248786",
  "text": "President Trump discussed immigration in his State of the Union address. #Migrant #DHS #BorderCare #ImmigrationUSA https://t.co/0vub0g0v0v",
  "truncated": true,
  "entities": {
    "hashtags": [
      {
        "text": "Immigration",
        "indices": [75, 82]
      }
    ],
    "urls": [
      {
        "url": "https://t.co/0vub0g0v0v",
        "expanded_url": "https://t.co/0vub0g0v0v",
        "display_url": "https://t.co/0vub0g0v0v",
        "indices": [117, 148]
      }
    ]
  },
  "metadata": {
    "source_lang": "en",
    "result_type": "recent"
  },
  "source": "<a href='\"http://www.facebook.com/twitter/\"' rel='\"nofollow\"'>Facebook</a>",
  "in_reply_to_status_id": null,
  "in_reply_to_status_id_str": null,
  "in_reply_to_user_id": null,
  "in_reply_to_user_id_str": null,
  "in_reply_to_screen_name": null,
  "user": {
    "id": 57482378,
    "id_str": "57482378",
    "name": "Reeves Immigration Law Group",
    "screen_name": "ReevesImm",
    "location": "LA, CA, US, Mexico",
    "description": "Call 800-799-0889 https://t.co/8zYw8uWw Immigration Experts Including Attorneys Formerly with the USCIS, State Department (US Embassy), and Immigration Courts.",
    "url": "https://t.co/8zYw8uWw",
    "entities": {
      "urls": [
        {
          "url": "https://t.co/8zYw8uWw",
          "expanded_url": "https://www.reeves.com",
          "display_url": "www.reeves.com",
          "indices": [8, 23]
        }
      ]
    },
    "description": {
      "urls": [
        {
          "url": "https://t.co/8zYw8uWw",
          "expanded_url": "https://www.reeves.com",
          "display_url": "www.reeves.com",
          "indices": [19, 42]
        }
      ]
    }
  },
  "protected": false,
  "retweets_count": 352,
  "retweets_count": 6,
  "likes_count": 3,
  "created_at": "Wed Jul 18 23:32:58 +0000 2009",
  "favorited_count": 6,
  "lang": "en",
  "time_zone": null,
  "geo_enabled": true,
  "verified": false,
  "statuses_count": 2648,
  "name": "Reeves",
  "contributors_enabled": false,
  "is_translator": false,
  "is_translation_enabled": false,
  "profile_background_color": "f0f0f0",
  "profile_background_image_url": "http://pbs.twimg.com/profile_images/799448740887408896/reeves_bg.gif",
  "profile_background_image_url_https": "https://pbs.twimg.com/profile_images/799448740887408896/reeves_bg.gif",
  "profile_image_url": "http://pbs.twimg.com/profile_images/799448740887408896/reeves_bg.jpg",
  "profile_image_url_https": "https://pbs.twimg.com/profile_images/799448740887408896/reeves_bg.jpg",
  "profile_banner_url": "https://pbs.twimg.com/profile_banners/57482378/1537268800",
  "profile_link_color": "000000",
  "profile_sidebar_border_color": "000000",
  "profile_sidebar_fill_color": "000000",
  "profile_text_color": "000000",
  "profile_use_background_image": false,
  "has_extended_profile": false,
  "default_profile": false,
  "default_profile_image": false,
  "following": false,
  "follow_request_sent": false,
  "notifications": false,
  "translator_type": "none"
},
  "geo": null,
  "coordinates": null,
  "place": null,
  "contributors": null,
  "is_quote_status": false,
  "retweet_count": 6,
  "favorite_count": 6,
  "favorited": false,
  "retweeted": false,
  "possibly_sensitive": false,
  "lang": "en"
}
```

# Sentiment Analysis

- Sentiment analysis analyzes the sentiment of text
- Uses words, capitalization and emojis
- Breaks words into three categories: positive, negative, neutral
- The compound gives the final overall sentiment of the text
- The average sentiment is what is used for the overall analysis

# Sentiment Analysis

- "I. LOVE. THIS. Speaker Nancy Pelosi jokes that "it will take days to fact-check" Trump 2019s State of the Union speech. "
  - 0.7865
- "President Trump discussed immigration in his State of the Union address. He is an idiot"
  - -0.5106

# Sentiment Analysis

- "President Trump discussed immigration in his State of the Union address. #Immigration #SOTU #GreenCard #Deportation2026 <https://t.co/GHvooCyNVV>"
  - 0.0
- "HATE President Trump discussed immigration in his State of the Union address. #Immigration #SOTU #GreenCard #Deportation2026 <https://t.co/GHvooCyNVV>"
  - -0.6633
- "Hate President Trump discussed immigration in his State of the Union address. #Immigration #SOTU #GreenCard #Deportation2026 <https://t.co/GHvooCyNVV>"
  - -0.5719
- "LOVE President Trump discussed immigration in his State of the Union address. #Immigration #SOTU #GreenCard #Deportation2026 <https://t.co/GHvooCyNVV>"
  - 0.7125
- "Love President Trump discussed immigration in his State of the Union address. #Immigration #SOTU #GreenCard #Deportation2026 <https://t.co/GHvooCyNVV>"
  - 0.6369



# Sentiment Analysis

- "HATE, anger, sad, mad"
  - -0.9375
- "HAPPY, love, good, great"
  - 0.9488

# Output

- With all this data, it can be a challenge to represent it
- Raw output of sentiment can be hard to read; lots of text & numbers
- Instead, use some form of visual tool; html, javascript
- Leaflet: Javascript package which is used to make maps similar to Google maps
- Highcharts: Javascript package for creating graphs of all kinds; used to create the time series line graphs
- <http://www2.cs.uregina.ca/~biezensa/index.html>