SWISS HAPPY MAPS

Kirtan Padh

Luis Medina

Tina Fang

Overview

- Data Preparation

- Machine learning (FastText)

- Visualization

Machine Learning: Text classification for Instagram tags

Why not image classification?

Negative picture

Neutral picture
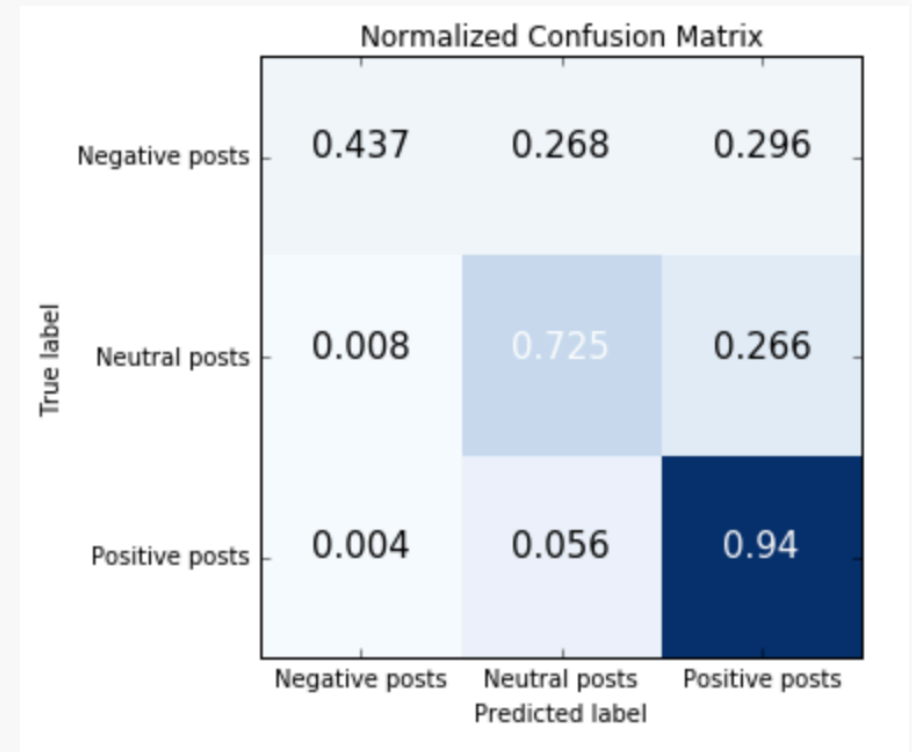
Why not image classification?

Positive picture

Positive picture

Text Classification: FastText

- Achieves really fast performance.

- Creates word representations which are then averaged into a text representation.

- No need to worry about the language, since each of the words will have its own weight related to the respective label.

Normalized Confusion Matrix

- **Number of Instagram posts in the training data set:** 1,504,278
  - 80% for training the model
  - 20% for testing the model

- Distribution of the sentiments in the training data set:
  - **Positive posts:** 71.98%
  - **Neutral posts:** 23.21%
  - **Negative posts:** 4.81%

Normalized Confusion Matrix

	Negative posts	Neutral posts	Positive posts
Negative posts	0.437	0.268	0.296
Neutral posts	0.008	0.725	0.266
Positive posts	0.004	0.056	0.94

True label / Predicted label

Results

- ■ **Number of Instagram posts in the test data set**: 6,845,983
- ■ **Precision of the created model**: 0.866590
  - *Positive posts:* 74.03%
  - *Neutral posts:* 23.9%
  - *Negative posts:* 2.07%
- ■ **Total of predictions**: 6,845,975
- ■ **Number of Instagram posts w/o prediction**: 8
- – *Bienna*
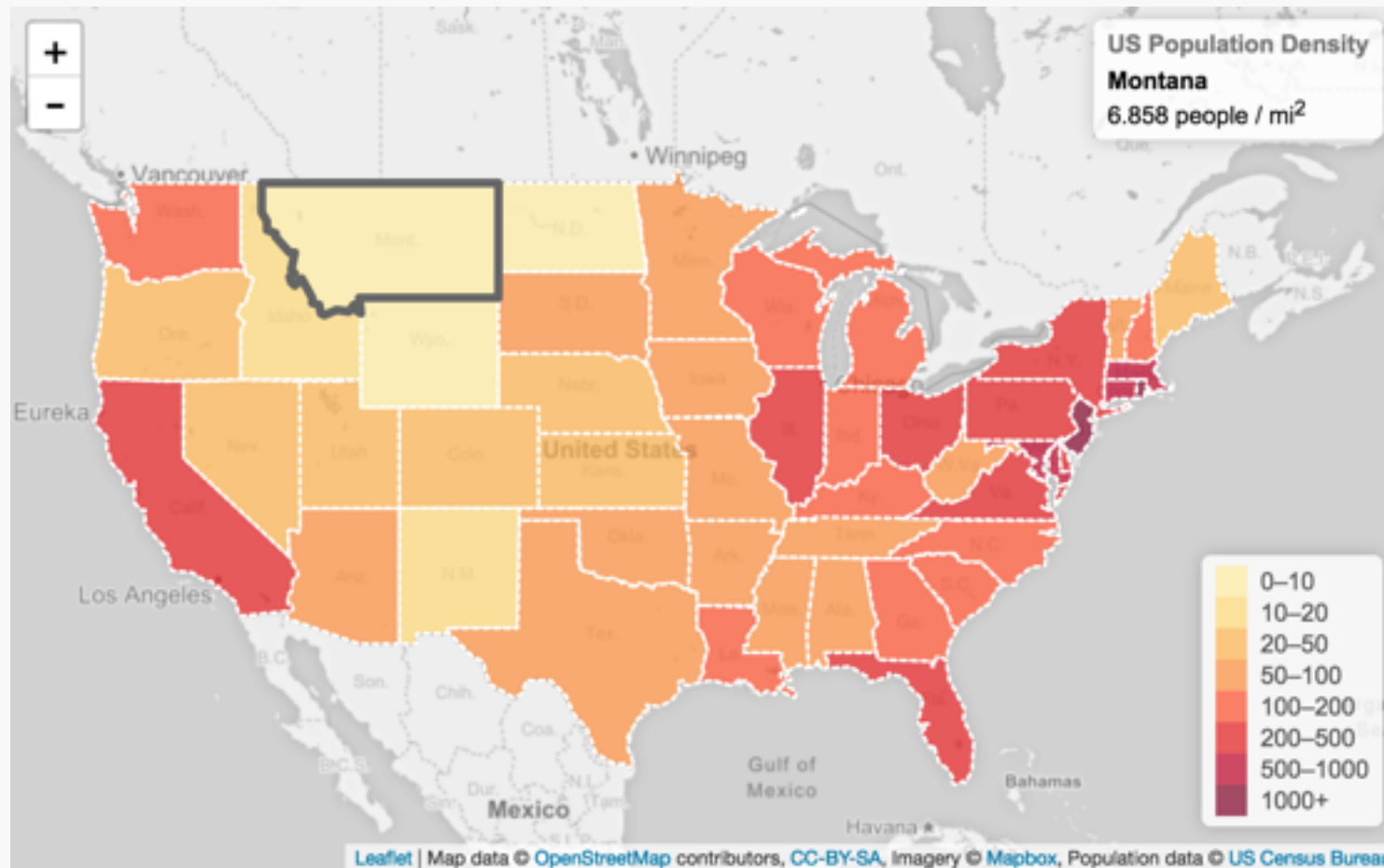- – *unterkulm*

Visualization

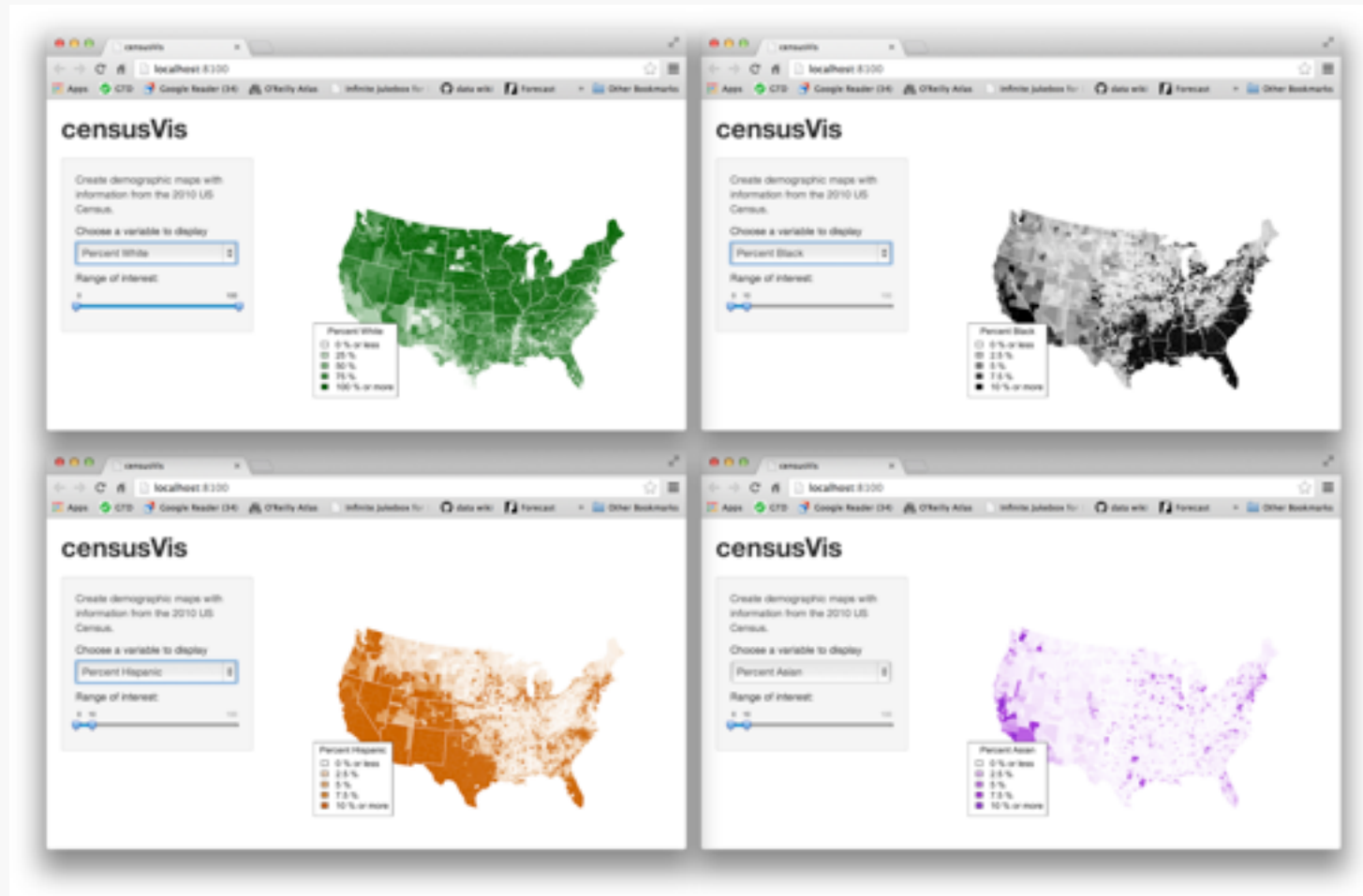10.5 Million tweets

3150 Filters

1 Map

Leaflet

- Javascript Mapping Library

- Dynamically hide and reveal information
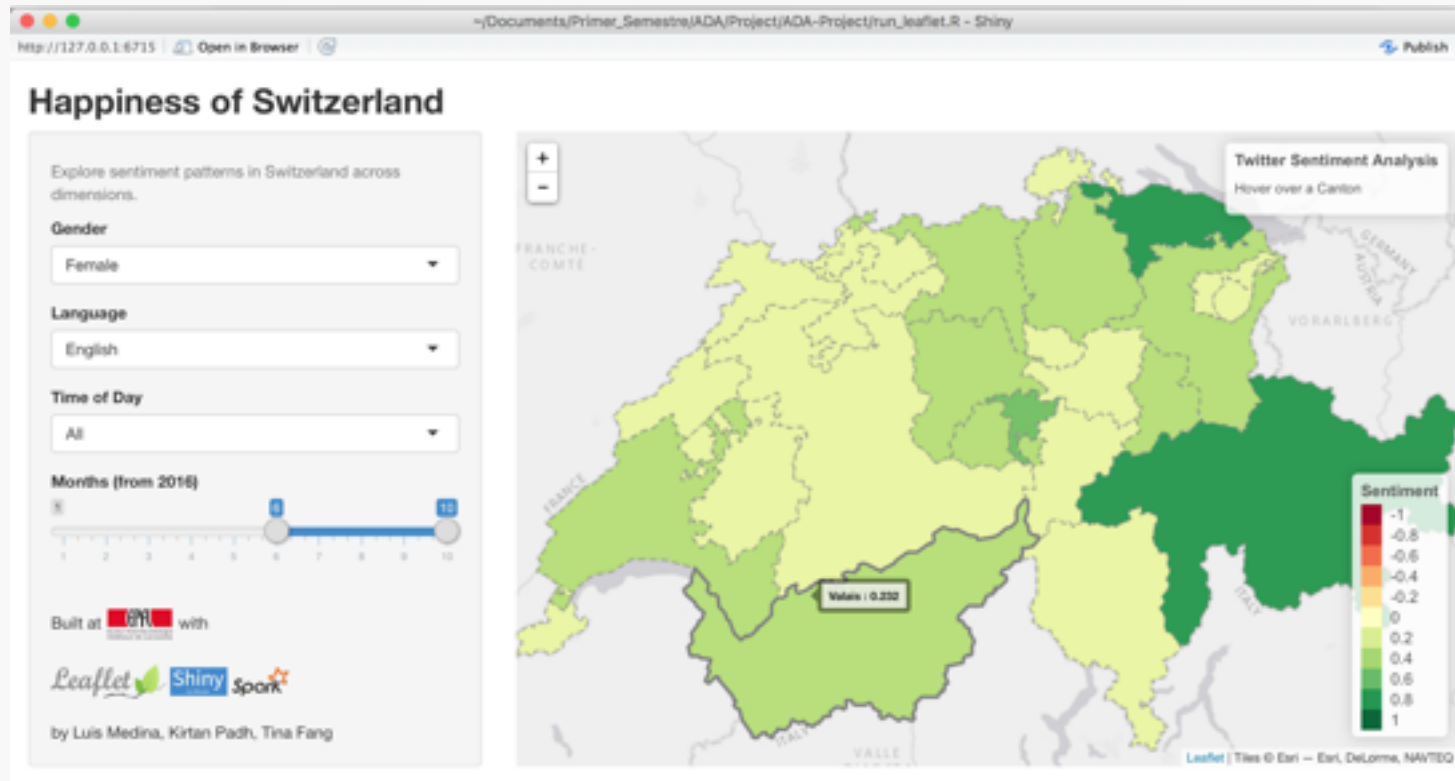
- Allows users to explore data

Other Tools: Folium, Vincent, Plot.ly, Bokeh, RMaps

Shiny

- Web Application Framework for R

- Dynamic filtering based on user selection

Result

- Demo

- Leaflet with Shiny

- Filter by:
  - Gender, Language
  - Time of Day
  - Month

- User Interactivity