DON'T DROWN, TURN AROUND

Detecting & Mapping Live Road Closures Using NLP and Modeling Techniques

Jasmine Vasandani
Phillip Dibert
Eric Kropf

When disasters and unexpected events cause road closures, navigation systems are slow to share updates with the public.

How might we detect and announce road closures as soon as they happen?

Methodology

Data Acquisition	Exploratory Data Analysis	Pre-Processing & Modeling	Activation
• Twitter • News API • 511 Alerts	• Natural Language Processing (NLP) • Statistical Analyses	• Classification Modeling • Soft Cosine	• Location Detection • GIS Mapping

OUR PROCESS


Data Acquisition
Exploratory Data Analysis
Pre-Processing & Modeling
Activation

Data Acquisition

Twitter

13k Tweets

All tweets from Minnesota official transportation accounts between April 2016-April 2019.

Minnesota 511

14 Live Road Closure Announcements

Live road closures verified by MN DOT.

News API

1k articles

News articles from across the world with relevant keywords (road, closure, etc.)

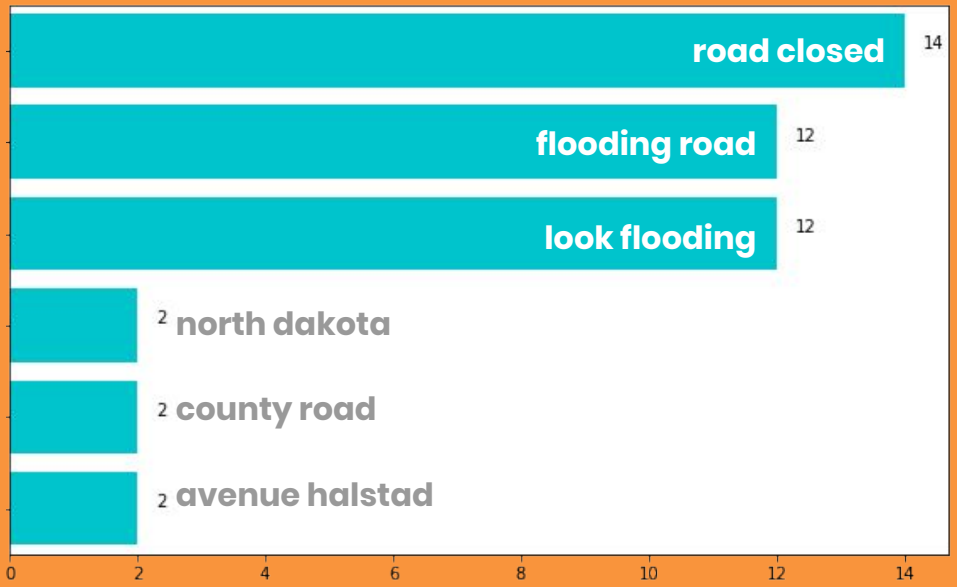
SUPERVISED MODEL

Twitter & MN511

Exploratory Data Analysis: MN511

Parameters for True Positives

Top 6 Bigrams from Alerts that Confirm Road Closures (MN511)



+ All road closures have the words “road closed”

+ Can use to filter for road closures in tweets

- Dataset too small

SUPERVISED MODEL

Twitter

Modeling: Tweets

Turning an Unsupervised Problem into a Supervised One

GOAL

Build a model that determines if a tweet is announcing a road closure

---

PROCESS

1. Using EDA from MN511 and tweet EDA, determine road closure tweets from 14k tweets dataset.
2. Identified 1,133 road closure tweets.
3. Create a balanced class dataset to train the model.

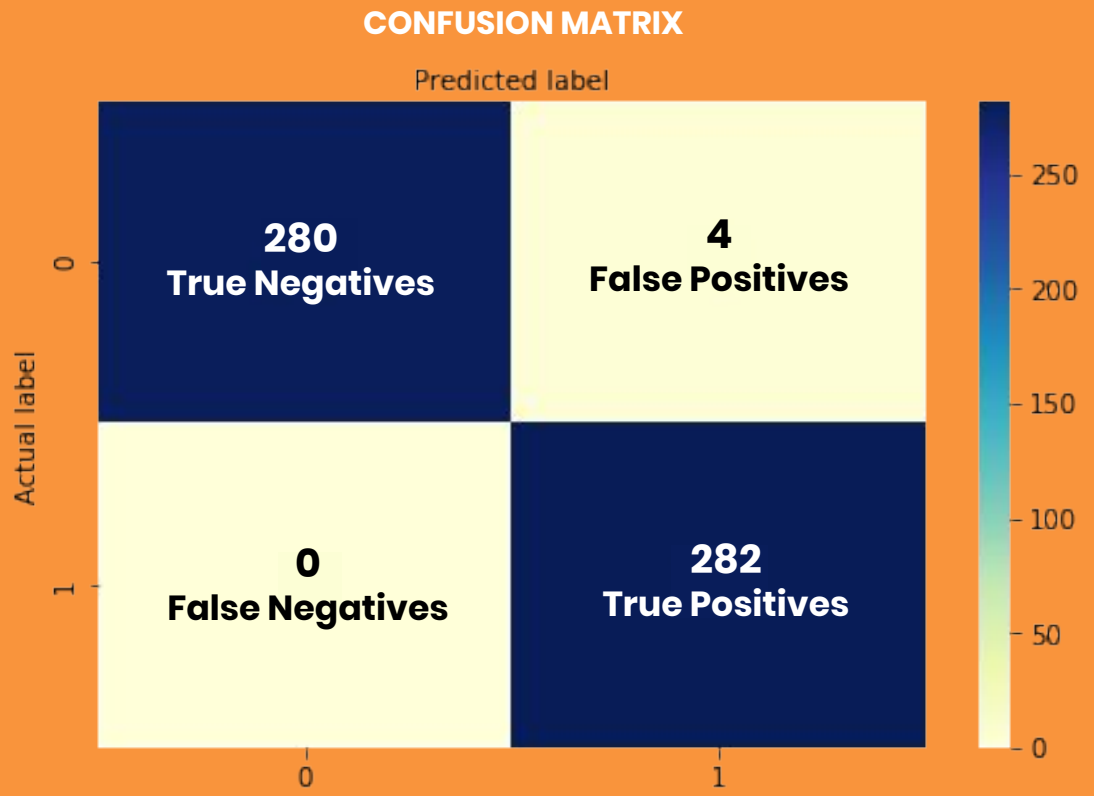
Modeling: Tweets

TARGET (Y)
Road closure: 1
Not road closure: 0

PREDICTOR (X)
Tweets!

BASELINE SCORE
50%

BEST MODEL
Logistic Regression
Count Vectorizer



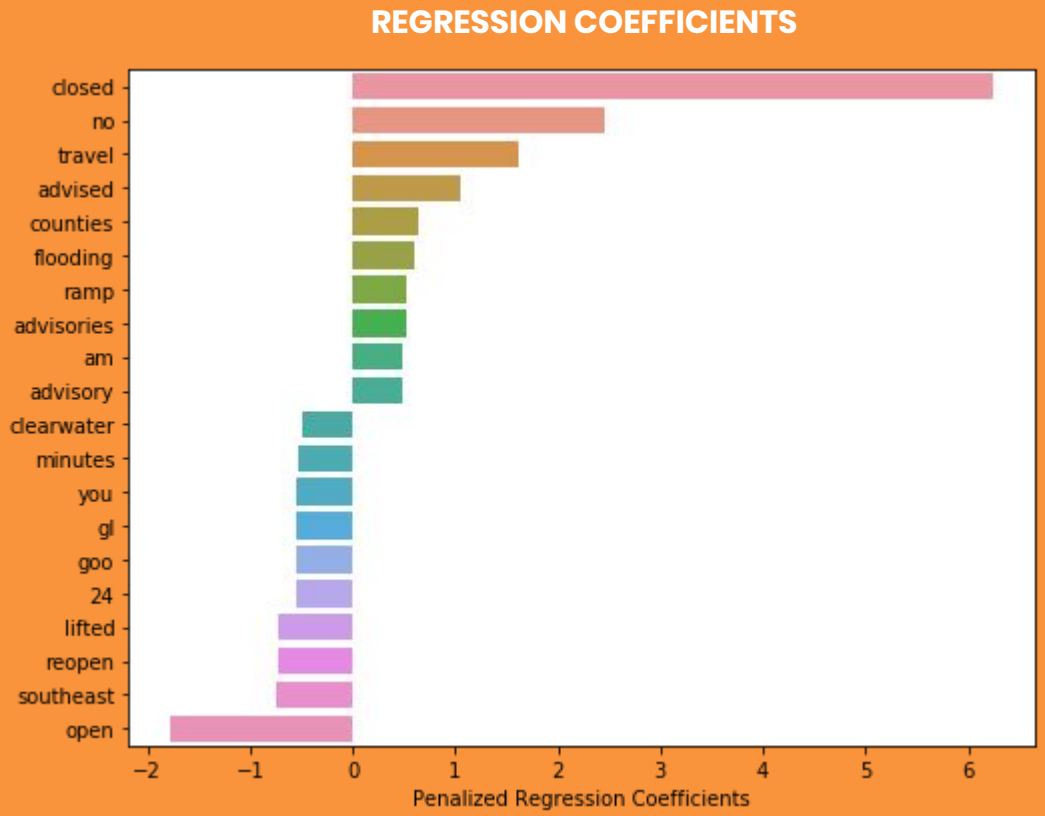
Modeling: Tweets

TRAIN SCORE
100%

TEST SCORE
99%

COEFFICIENT ANALYSIS

- "closed" 514.91 times likely to be a road closure.
- "open" 5.92 times likely to be a non-road closure.



Running Model on 126 Unseen Tweets

TRUE POSITIVES

- Annapolis St is closed to thru traffic btwn Smith Ave and Cherokee Heights Blvd
- Hwy 60 at Wabasha at railroad bridge near Hwy 61 is closed because of high water on the road.

TRUE NEGATIVES

- ROAD CLOSING SOON - Hwy 93 between Hwy 169 and Le Sueur will be closing soon.
- SB 169 St Peter to Mankato is now open!

FALSE POSITIVES

- SB Hwy 61 in Cottage Grove: Right lane closed between Jamaica Ave and Innovation Rd tomorrow (4/16) from 9:30am to 2:30pm.

FALSE NEGATIVES

- Hwy 169 - BLOCKED - due to mudslides and traffic. Please find alternate route!!!!
- High water still causing closures in NW Mn: Hwy 1 east and west of Oslo Hwy 2B

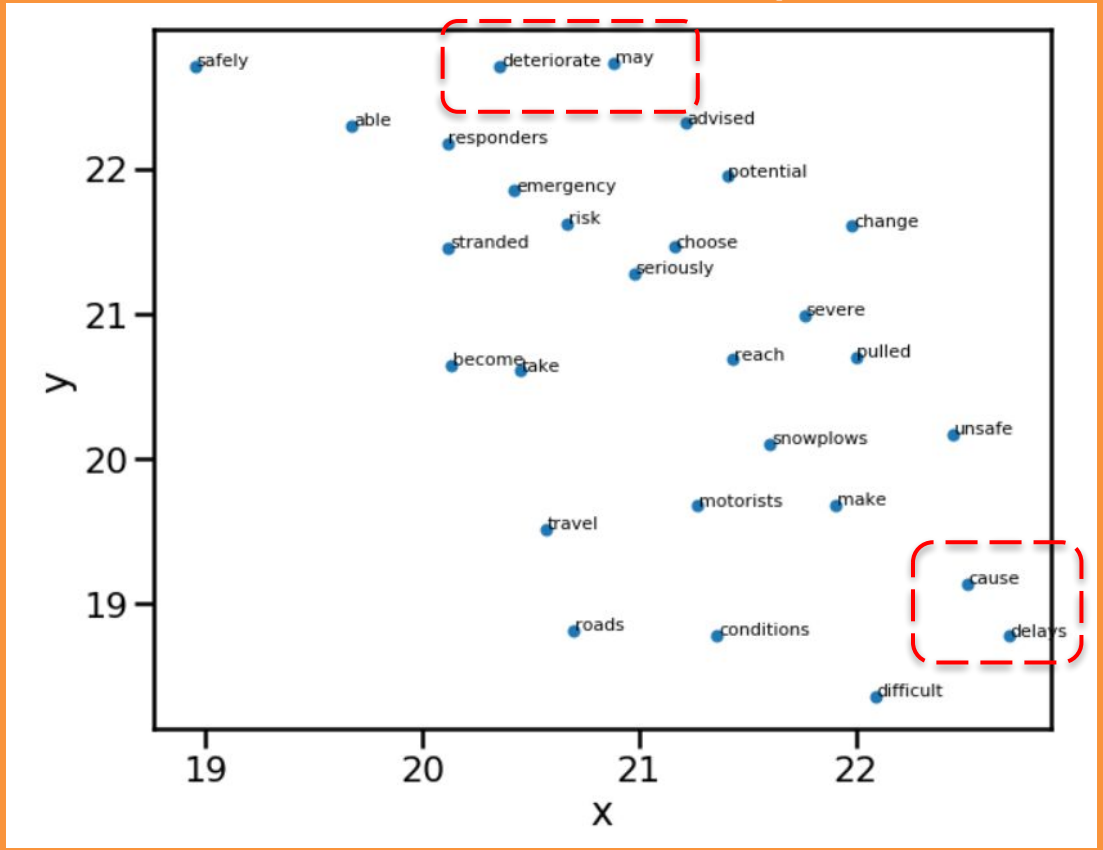
UNSUPERVISED MODEL

News

Exploratory Data Analysis: ND-511

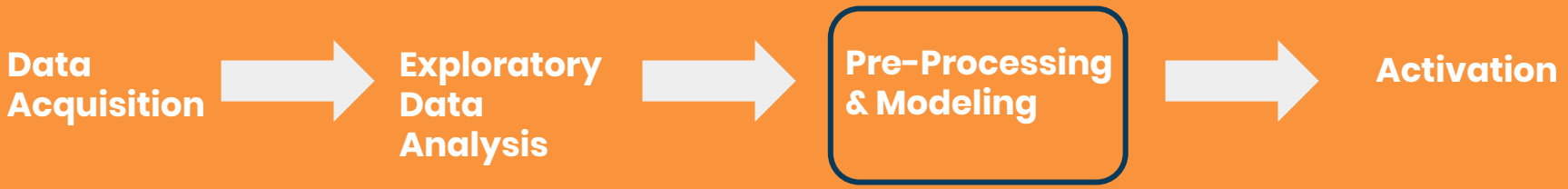
- Apply t-SNE to vectors from ND-511 to visualize relationships

- Related words appear close to each other in the 2D rendering



UNSUPERVISED MODEL

News



Modeling: News API

Unsupervised NLP model

GOAL

Build a model that determines if a news article is announcing a road closure

PROCESS

1. Vectorize ND511 announcements and news articles
2. Remove place names (SpaCy) and stop-words
3. Calculate soft cosine similarity between announcements and news
4. High soft cosine similarity indicates that news article is likely to be announcing a road closure

UNSUPERVISED MODEL

News

Not all articles with high soft cosine values are announcing road closures

FAST COMPANY

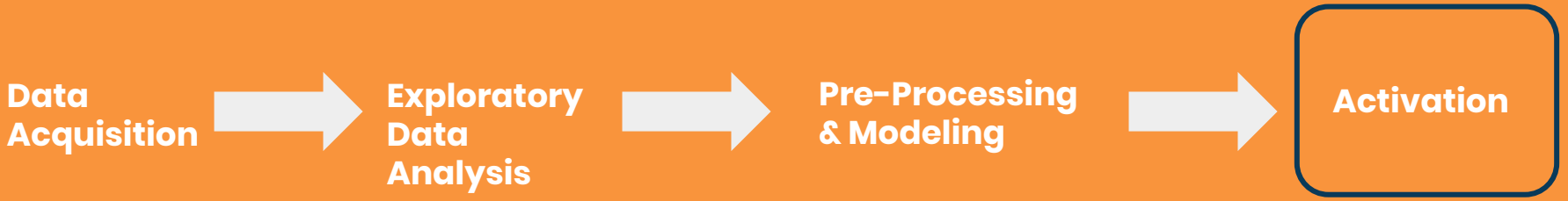
By the middle of the century, as climate change progresses, extreme flooding is likely to increase across some southeastern U.S. states. AT&T, which has to deal with the immediate aftermath of any storm when phone service goes out, now knows where flooding is more likely–down to the level of each neighborhood.

The Washington Post
Democracy Dies in Darkness

Gridlock

George Washington Parkway's northbound lanes to be closed through the weekend after sinkhole forms

ACTIVATION
Post-Modeling Process



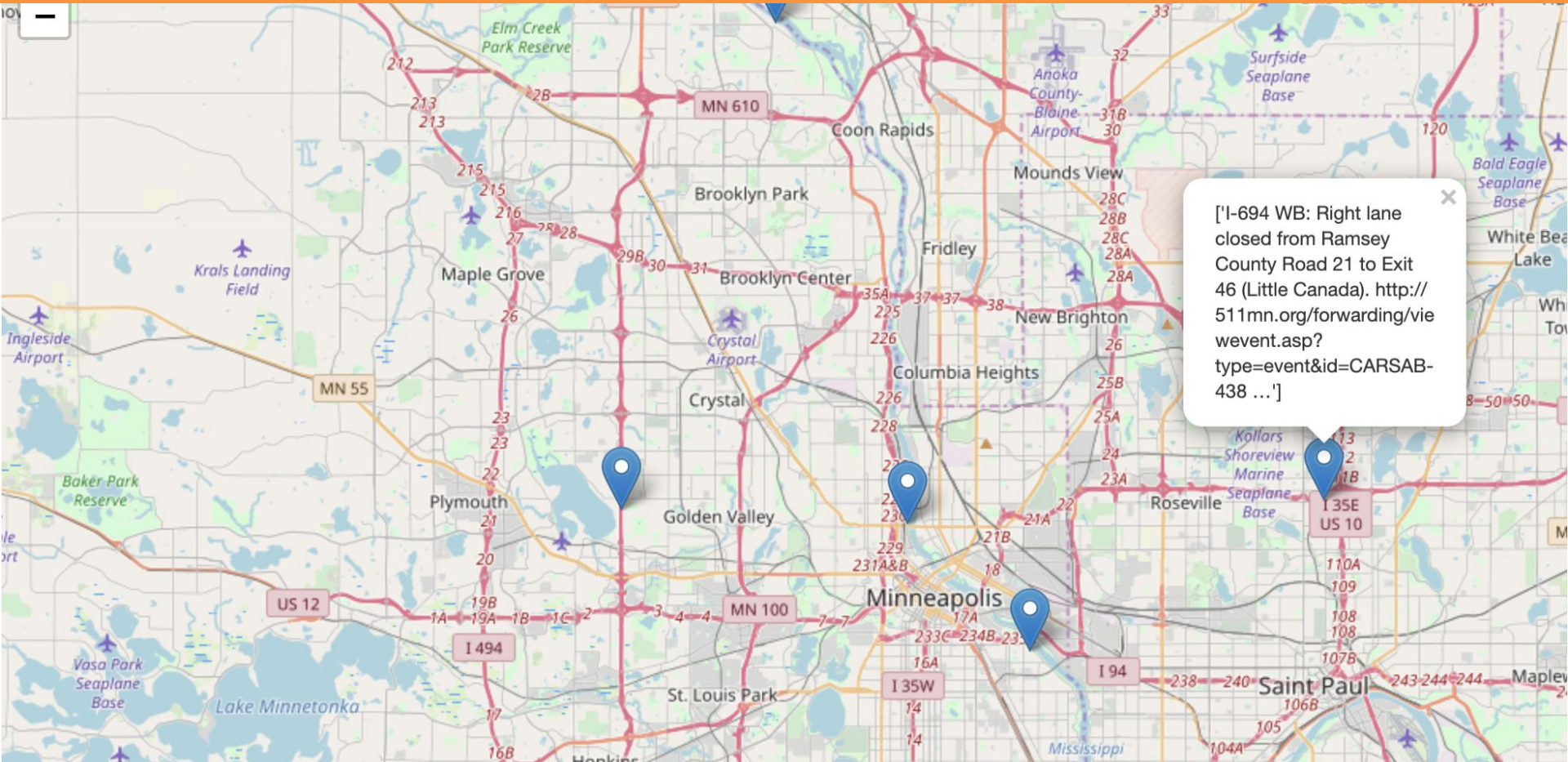
Extracting Location from Text w/ NO Geotag

Get to know your tweets—EDA is vital!

With the modifications below, increased SpaCy's ability to detect location by **98%**:

- Replace common prepositions with "at"
- Spell out location abbreviations (e.g., "Hwy", "NB", "SB", etc)
- Remove punctuation

Mapping Live Road Closures



ZOOMED OUT

ZOOMED IN

['Hwy 169 Plymouth: Ramp from Medicine Lake Road to SB 169 closed 7/12 thru mid-Aug. http://ow.ly/nP5J30dy9ab']

Kilmer Lane North
Mendelssohn Avenue North
East Medicine Lake Boulevard
23rd Avenue North
English Circle
Stroden Circle

Lessons Learned

+ Getting creative with modeling works
+ Overall, the process is actually simple

- Data acquisition is a lengthy process
- Needed more data on road closures to build an accurate model
- Not all sources give precise road closure information

Future Research

- Continue to refine models by feeding additional relevant articles and Tweets
- Test out model on different geographies