A sentiment analysis of the 2019 Australian election

• • •

Jack Carlisle
Mohammed Karaki
Cristian Rodriguez

Sentiment Analysis

- Classify text as positively valenced or negatively valenced.
- Modern ML techniques allow us to apply sentiment analysis to large datasets.
- Apply these techniques to a political context, namely the 2019 Australian election.

Model eliminates irrelevant data, identifies strongly valenced words and phrases.

Model assigns sentiment value to tweet based on presence of strongly valenced words and phrases.

"Had so much fun at the beach with my best friends Jill and Demetrius!"

"~~Had so much~~ fun ~~at the beach with my~~ best friends ~~Jill and Demetrius~~!"
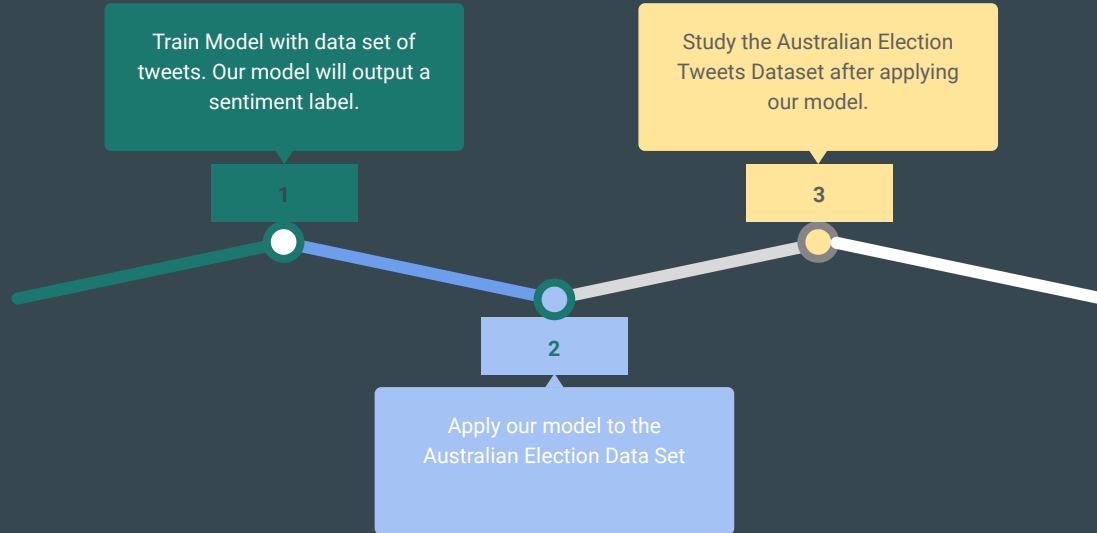
4

"Ugh, I hate waiting in line at the DMV!"

"Ugh~~, I~~ hate ~~waiting in line at the DMV!~~"

0

The Process

Train Model with data set of tweets. Our model will output a sentiment label.

1

Apply our model to the Australian Election Data Set

2

Study the Australian Election Tweets Dataset after applying our model.

3

The Training Dataset

- Dataset consists of 200 thousand tweets.
- Each tweet is pre-assigned a valence (0 = negative, or 4 = positive).
- Tweets were randomly sampled, in that their content is not about any particular topic.
- We use this data set to train our ML models, and test performance.

Training our Model
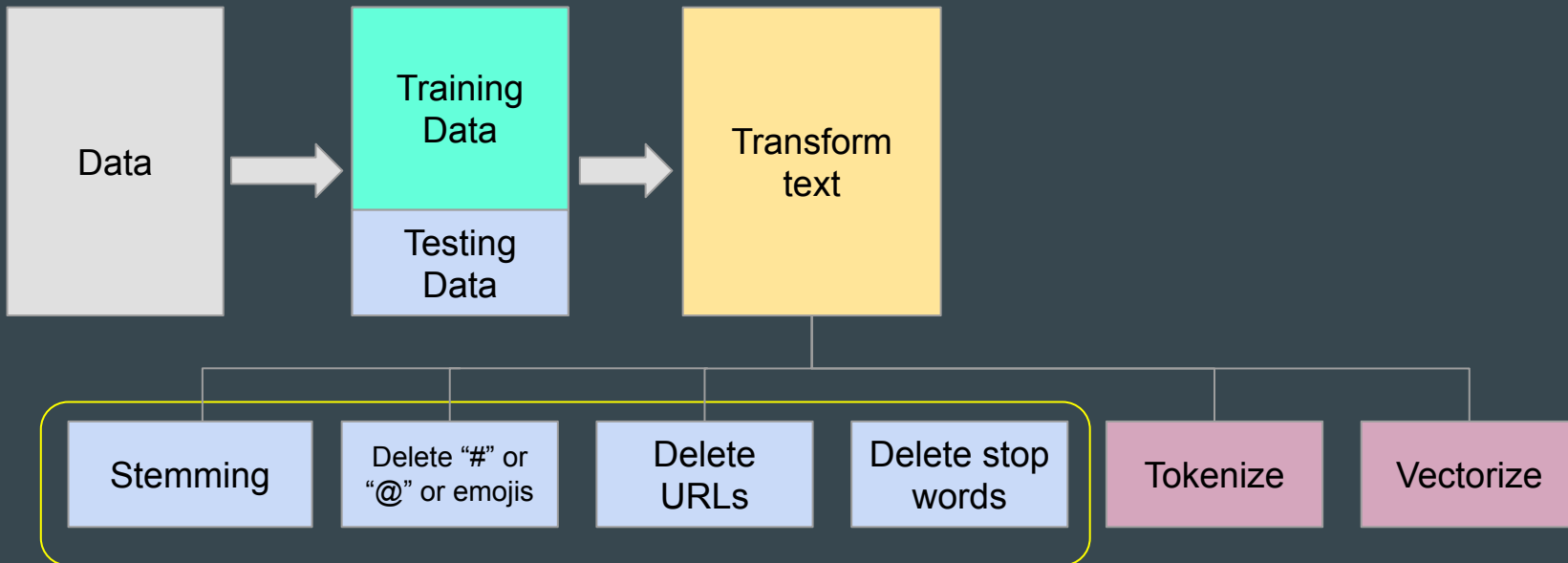
Training our Model
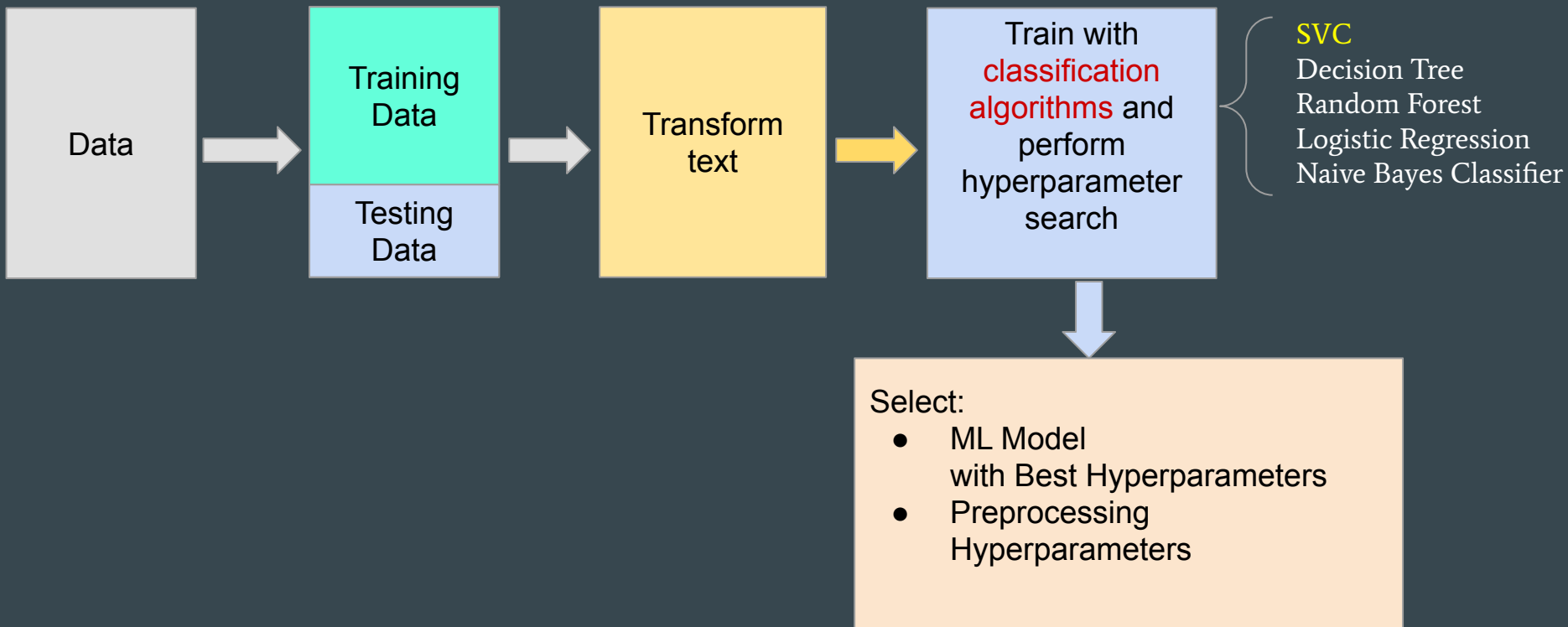
Some Training Results Examples using SVC

Remove URLs	✗	✗	✗	✗	✅	✅
Remove "@"	✗	✗	✗	✗	✗	✗
Remove "#"	✗	✗	✗	✗	✗	✗
Remove Emojis	✗	✗	✗	✗	✅	✅
Stemming	✗	✗	✅	✅	✗	✗
Remove StopWords	✗	✅	✗	✅	✗	✅
Best obtained Accuracy among different ML-Models	0.80690	0.783619	0.807575	0.782231	0.806537	0.782134

Model Selection

The Best Model with 80.75% accuracy was

   Linear Support Vector Machine Classifier

   using 2-grams

   without inverse-document frequency reweighting (TD-IDF)

   Preprocessing text with stemming and not removing stop-words, #s, @s, URLs or emojis.

Confussion Matrix Linear SVC

| Actual	negative	positive
negative	26877	6244
positive	6430	26449

Predicted
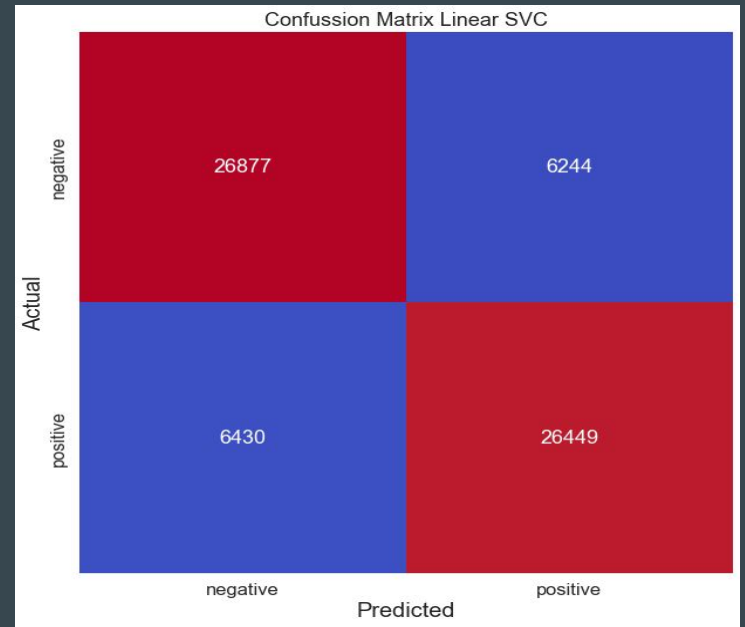
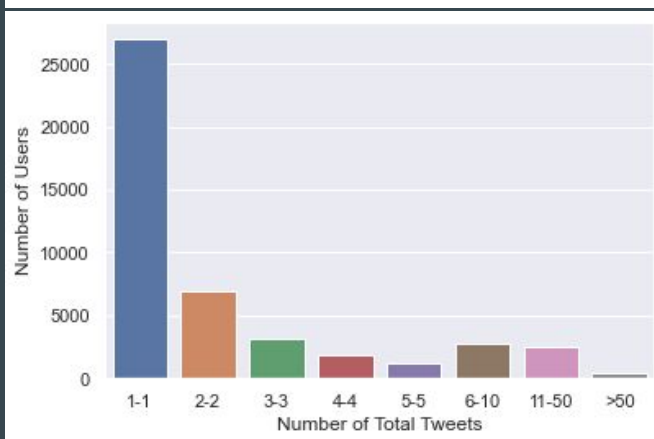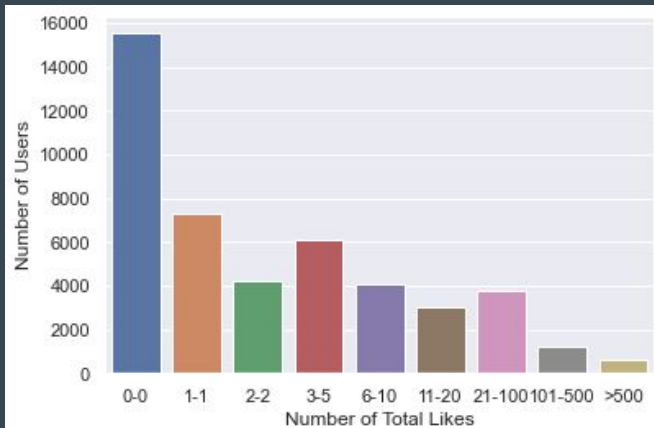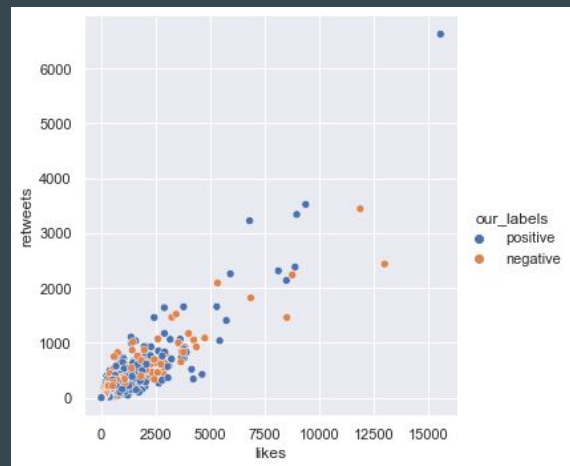# The 2019 Australian Election

- Primary competitors include "the Coalition" (conservative party), and "Labor", (progressive party).
- Opinion polls suggested that Labor would win a majority.
- Surprisingly, the Coalition maintained its majority.
- How can our chosen ML model explain this outcome?
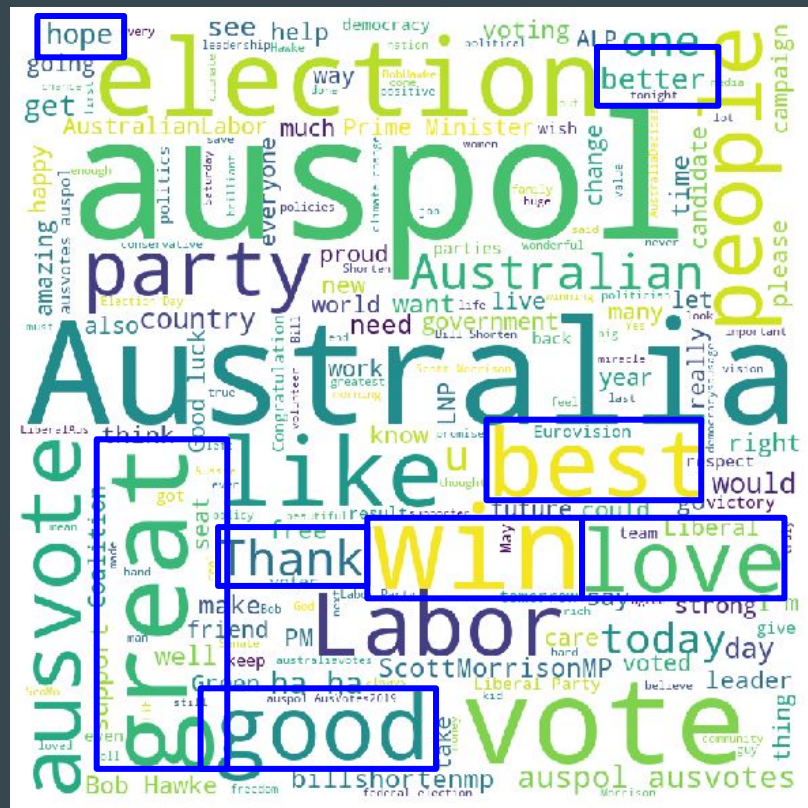
Applying our Model to the 2019 Australian Tweets Dataset

Understanding the dataset:

- There are 183 380 after deleting some repetition
- Written by 45 862 distinct users
- The User with Most tweets was 💧 Sur Plus with 937 tweets
- The tweets were between 2019-05-10 until 2019-05-20
  and the **election results** became available at **2019-05-18 21:30 PM**
- The average number of likes was 11.14
- The average number of retweets was 3.81

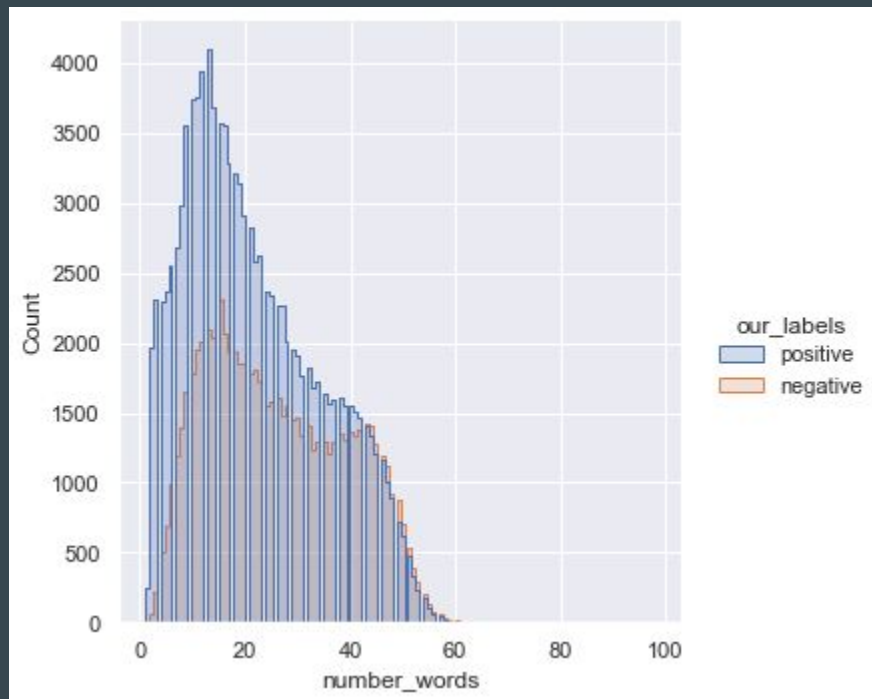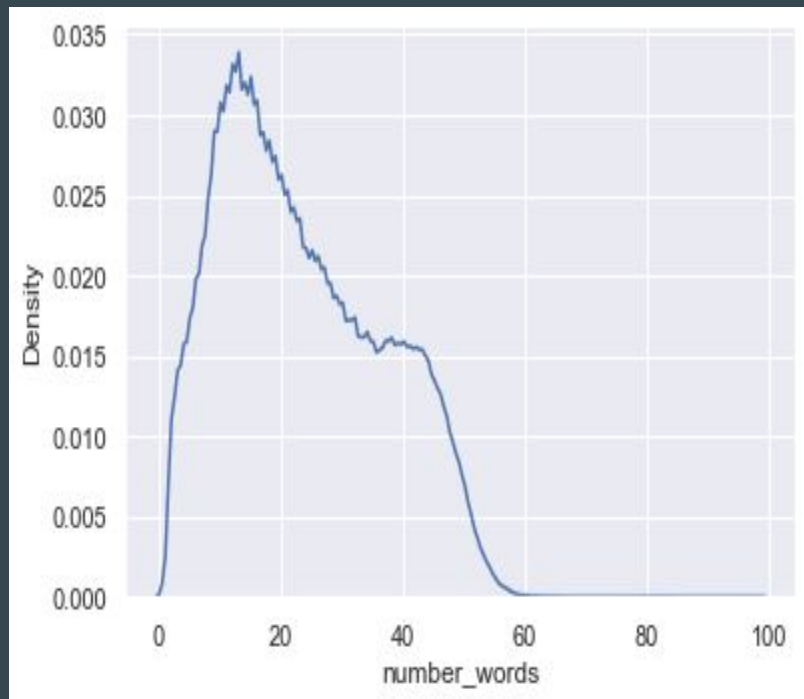# Applying our Model to the 2019 Australian Tweets Dataset
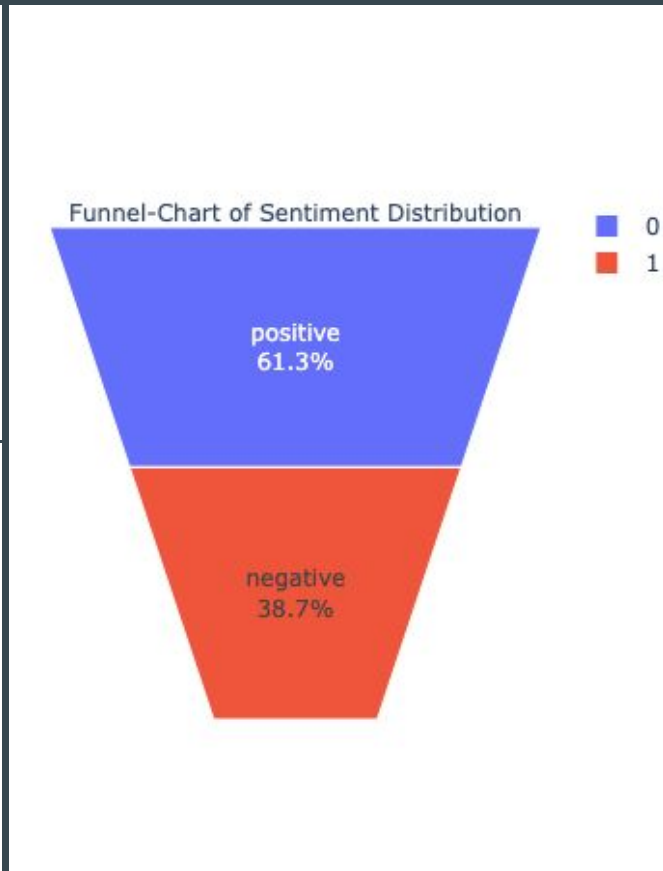
Positively Valenced Tweets — Negatively Valenced Tweets

Applying our Model to the 2019 Australian Tweets Dataset

Applications of our model

- Our model could be used to **predict party affiliation** based on tweet sentiment over time.
- Our ability to implement this application is limited since party affiliation for our dataset is unknown.
- We manually pulled a small sample of tweets from our dataset, the party affiliation of whose user is known. Our predictive model had an **accuracy of 80%** on this sample.
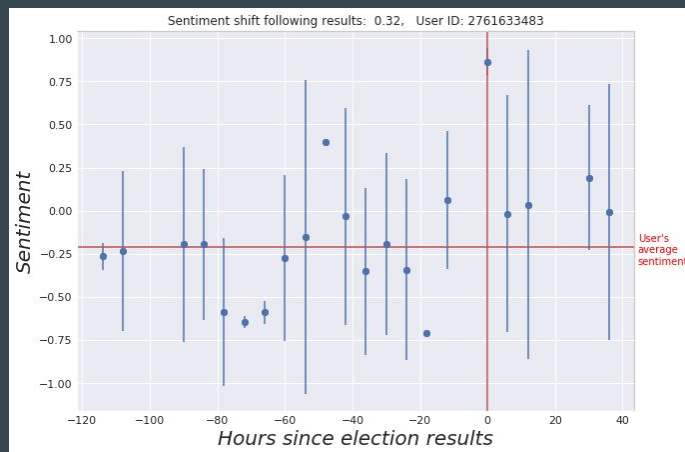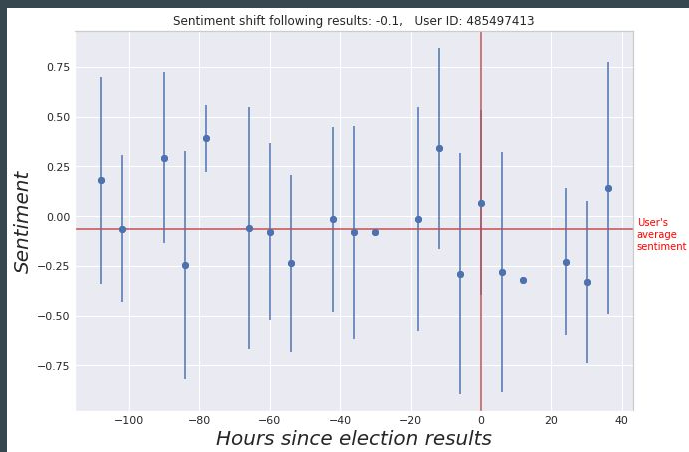
Conclusion

- In the political sphere, understanding personal opinion on a large scale is crucial to political campaigns, activists, and the news media.

- The abundance of online data combined with modern ML techniques provide valuable tools to obtain such insights.