

Online Community Monitoring

Tracking Changes in Subreddit Sentiments

Natural Language Processing

DATS 6312, Section 11

Team 3, Final Project

Ashish Telukunta, Jonathan Schild, Thomas Stanton

Project Outline

1. Conceptualization/description
2. Training model to predict emotion and sentiment of text
3. Scrape Reddit data
4. Apply model(s)
5. Plot sentiments over time

Model Approach

1. Data
2. Model identification
 - a) Positive/Negative sentiment
 - i. Hugging Face's pipeline
 - b) Emotions
 - i. BERT1, BERT2 (pseudo-labeling)
 - ii. RoBERTa
 - iii. ELECTRA
3. Tuning
 - a) Hyperparameters
 - b) Optimizer, Loss function

Model Performances

Model	Validation Accuracy	F1 Score
BERT1 (w/o Reddit data)	0.9520	0.9582
BERT2 (w/Reddit data)	0.9100	0.9140
Electra	0.9531	0.9530
Roberta	0.9502	0.9502

Reddit Scraper

1. Praw
2. Duration restrictions: day, week, month, year
 - a) No ability to specify date ranges
3. Data retrieved
 - a) Posts & comments; specify number of each
 - b) Dates; enable grouping and averaging

Sentiment Logging Training

Data Sources

Pre-Loaded Data [Live Data](#)

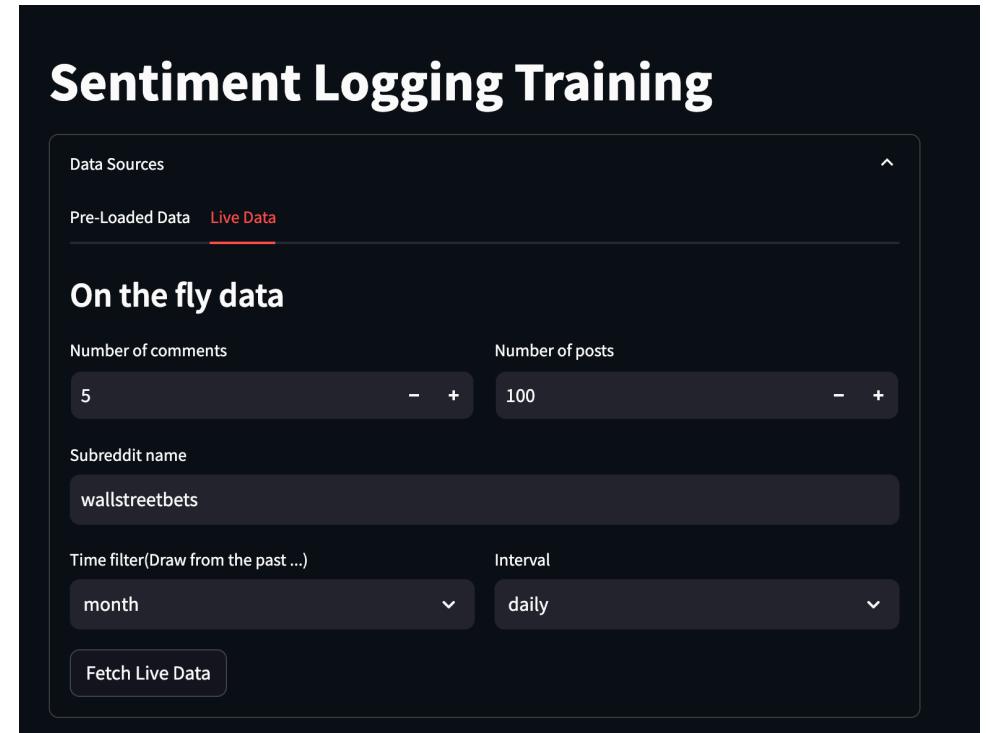
On the fly data

Number of comments: 5 Number of posts: 100

Subreddit name: wallstreetbets

Time filter(Draw from the past ...): month Interval: daily

[Fetch Live Data](#)

A screenshot of a web-based application titled "Sentiment Logging Training". The interface is dark-themed. At the top, there's a header "Data Sources" with tabs for "Pre-Loaded Data" and "Live Data", where "Live Data" is currently selected. Below this is a section titled "On the fly data" containing input fields for "Number of comments" (set to 5) and "Number of posts" (set to 100). There are plus and minus buttons next to these input fields. Underneath is a field labeled "Subreddit name" with the value "wallstreetbets". Further down are "Time filter" and "Interval" dropdown menus, both currently set to "month" and "daily" respectively. At the bottom of the form is a button labeled "Fetch Live Data".

Model Application: Preview

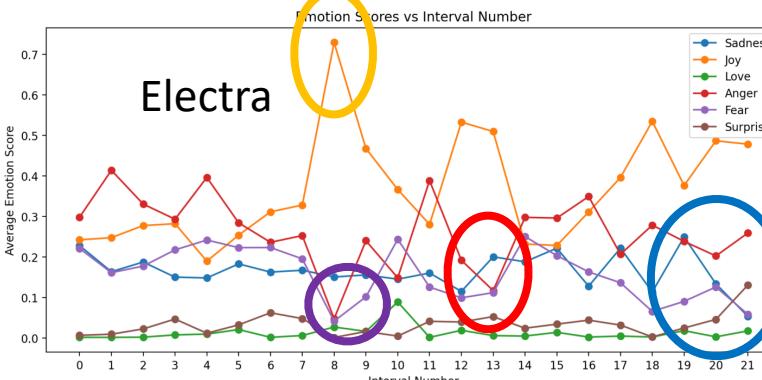
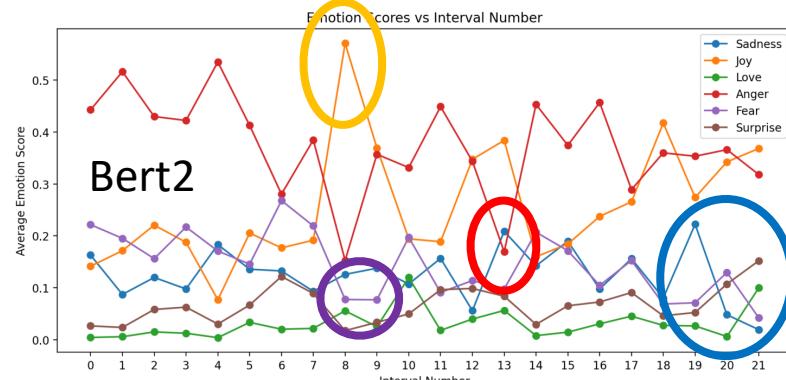
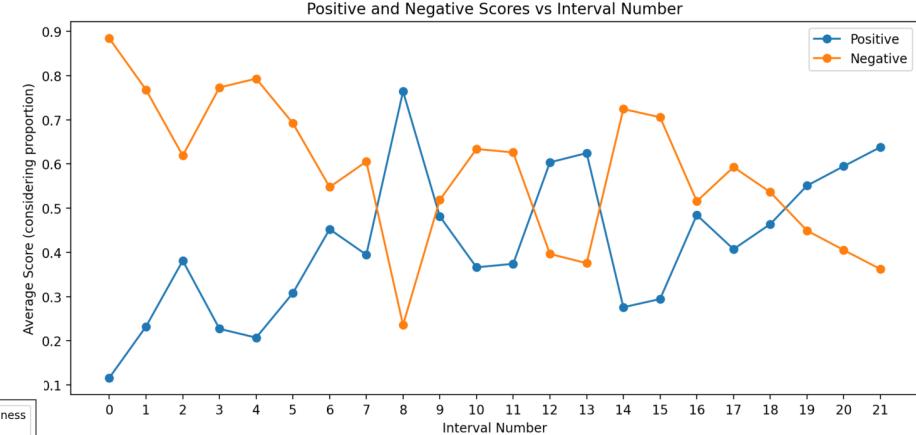
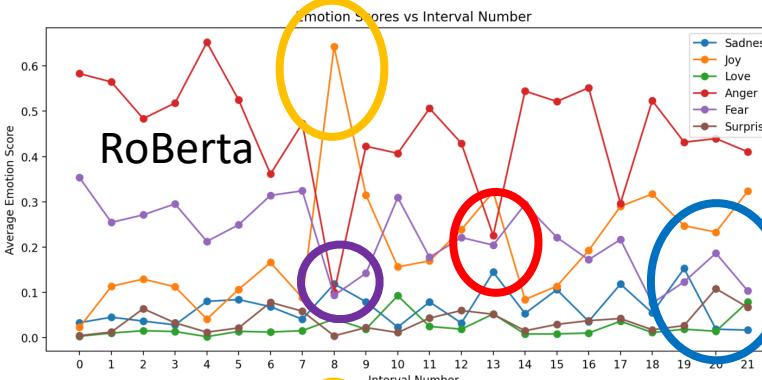
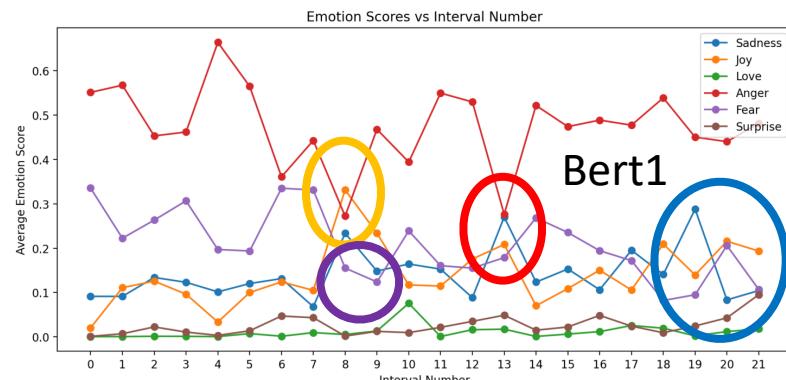
1. Scrape parameters

- a) Top 40 comments for the top 50 posts
- b) For past month, grouped to daily intervals

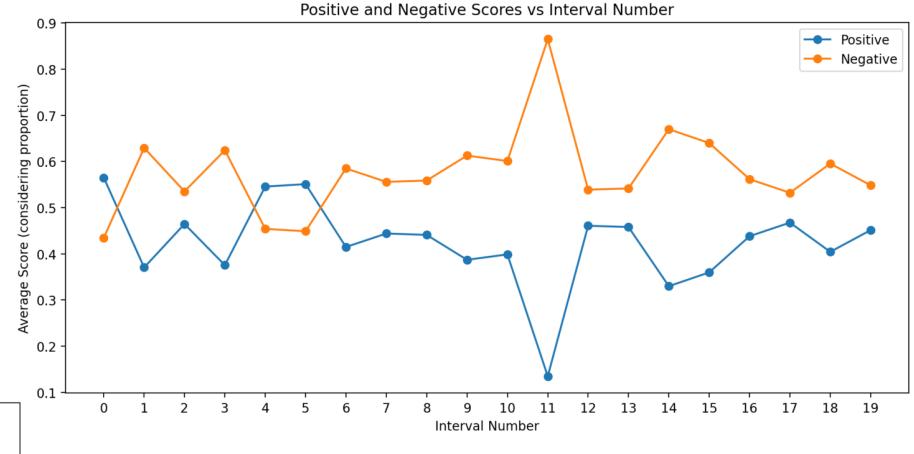
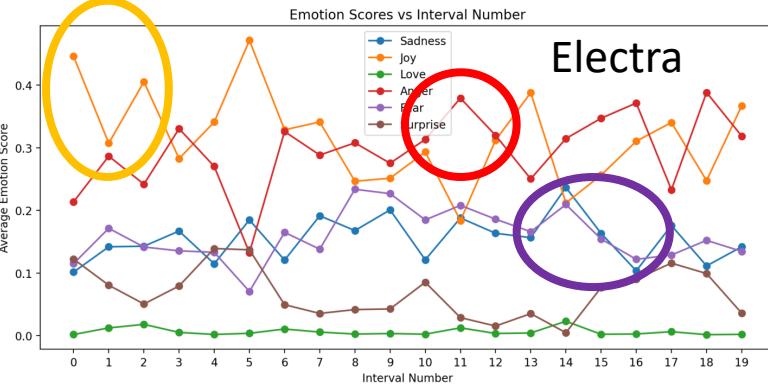
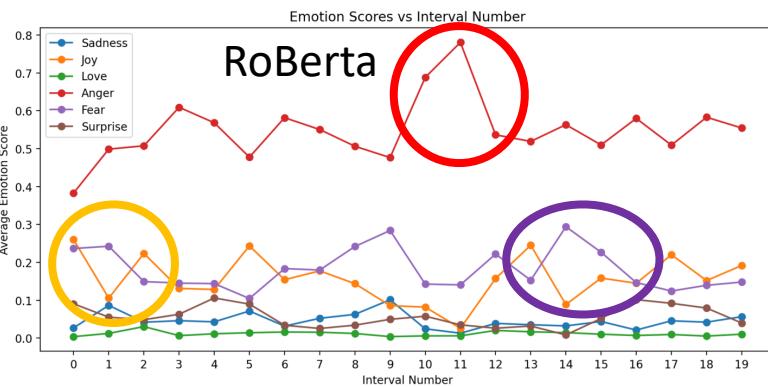
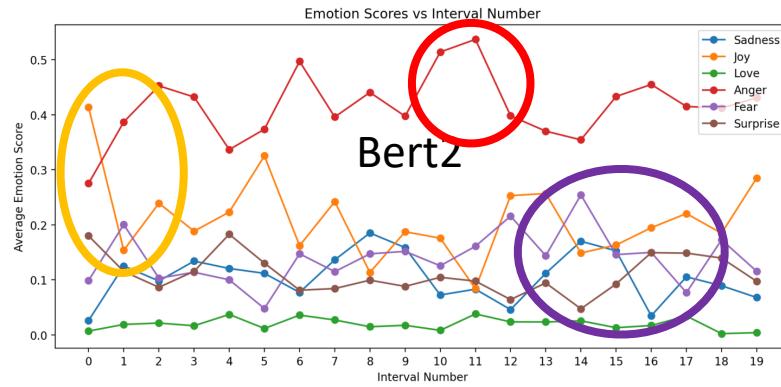
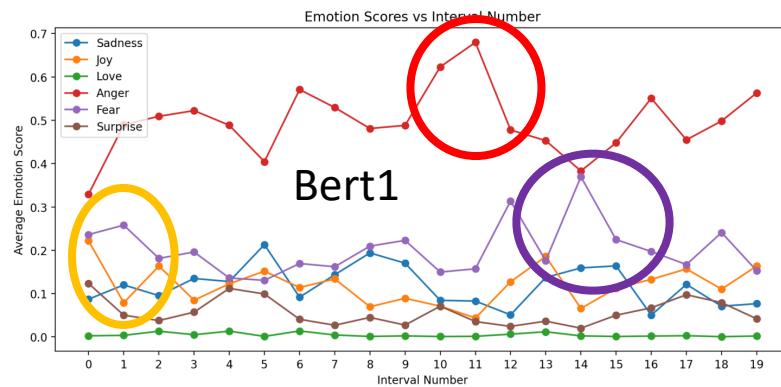
2. Examples...

- a) r/pics
- b) r/funny
- c) r/wallstreetbets
- d) r/gwu

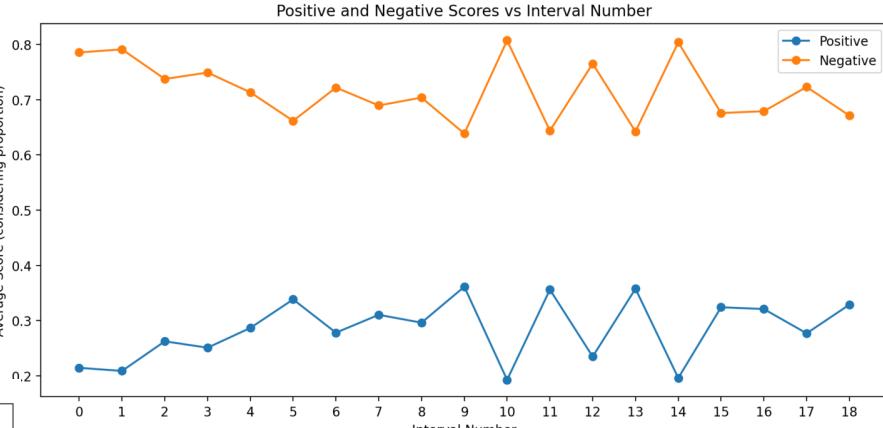
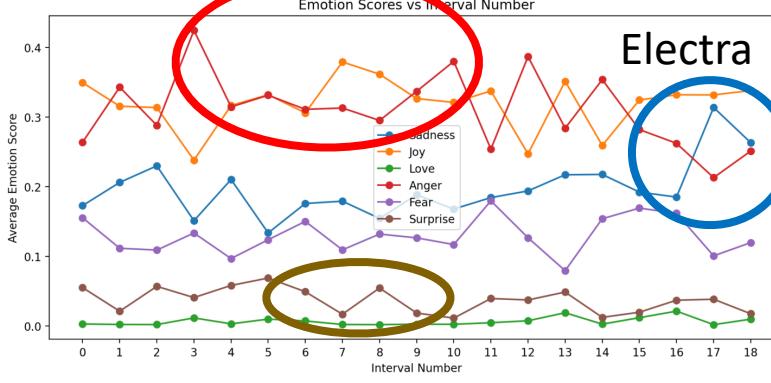
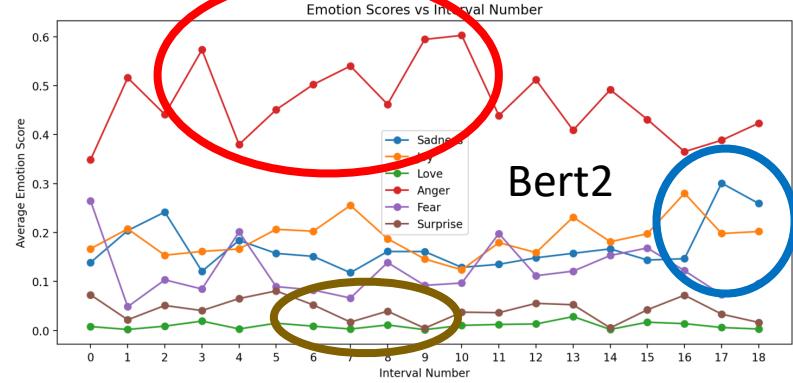
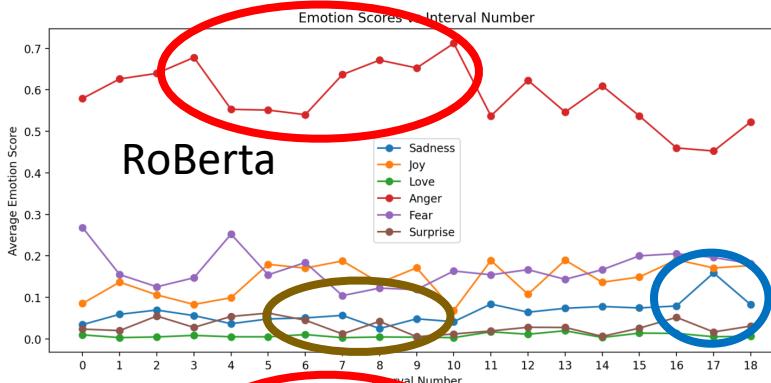
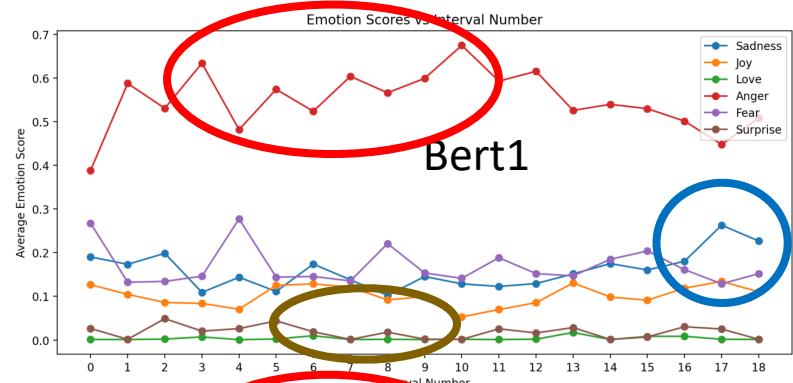
r/pics...



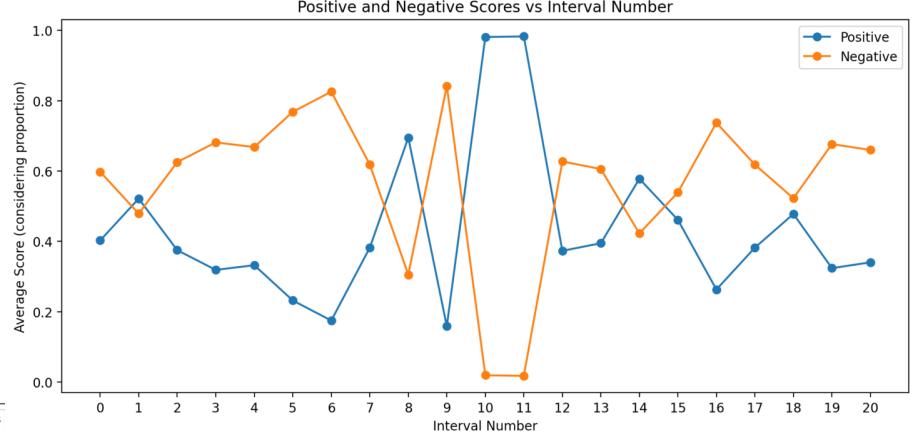
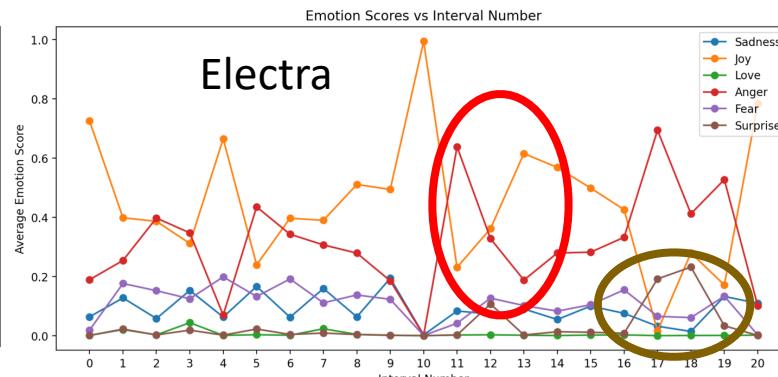
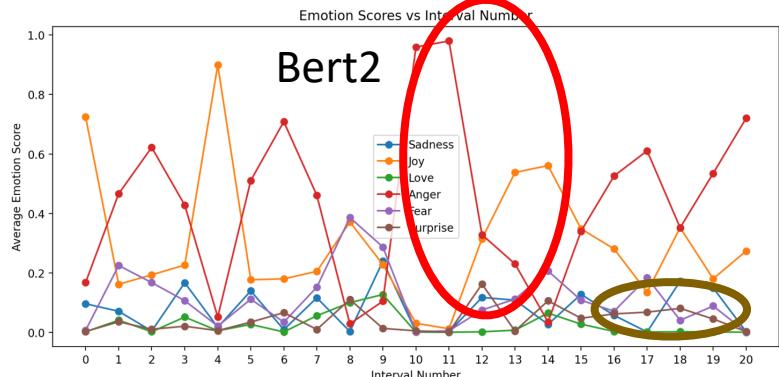
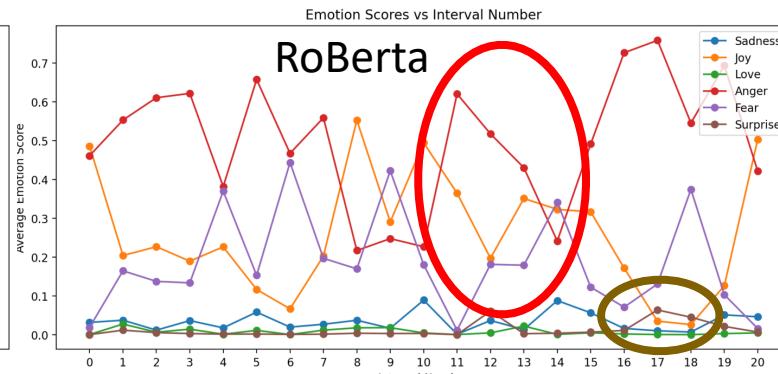
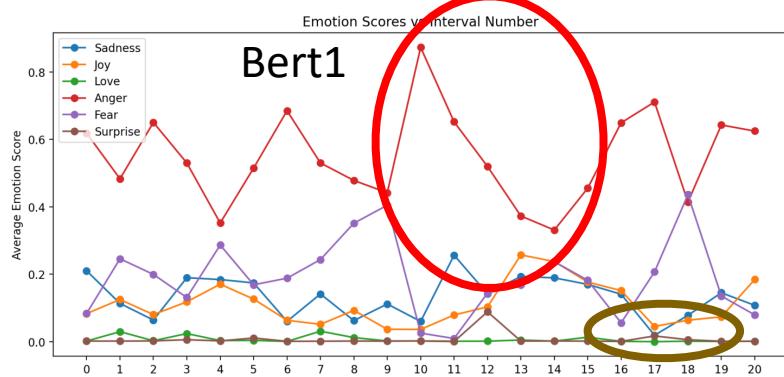
r/funny...



r/wallstreetbets...



r/gwu...



Demonstration...