Evolution of Hate Speech on Twitter since Elon Musk 's Acquisition

Focus on hate speech towards the Black community

04/13/2023

The team

Helena Caramelle

Sciences Po & HEC
Major: Public Policy

Shengyi Zhang

Sciences Po & Hertie
Major: Public Policy
-Global health stream

Georgios Saitis

Sciences Po
Major: Public policy
-Politics stream

Seng Moon Ja

MA Economic Policy
(CEU)

Context and background

- Elon Musk has been described as a "Free speech absolutist"
- Layoff of content moderation staff
- Reinstatement of right-wing accounts
- Multiple studies found that hate speech has <u>increased</u> on Twitter
- But Elon Musk has been claiming that hate speech has seen a <u>decline</u> on the social media platform

"Hate Speech Spikes on Twitter After Elon Musk Acquires the Platform"
(Benton et al., 2022)

Tweets from 10/22 to 10/28 collected from the Tweetbinder analytics program
"n-word," "k-word," "f-word," etc. examined to see if frequency of use increased in correlation to the change in ownership and stated platform focus.

Results:
- "Immediate spike in Tweets using one or more of the analyzed hate terms in the period leading up to Musk taking over the company"
- "Sentiment analysis validated numeric assessment, with 67.2% of tweets using these terms found to have a negative tone"
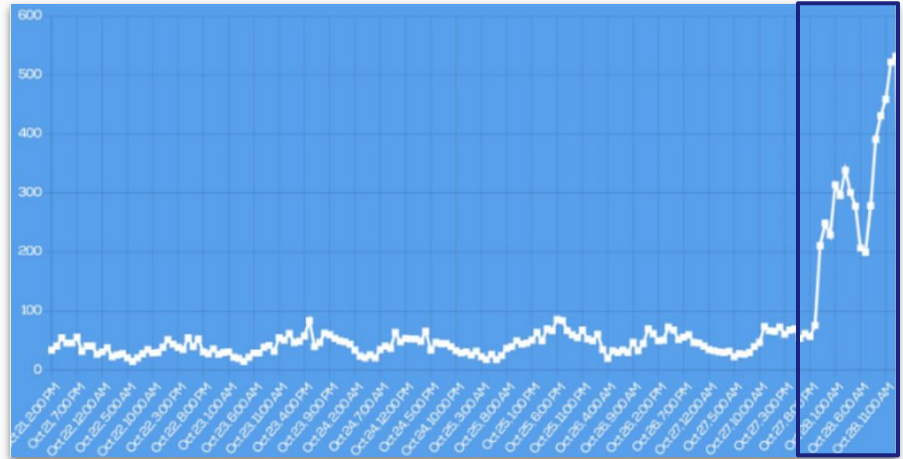
Preliminary research (2/3)

Tweets from 10/31 to 11/06 (first full week under Musk's ownership)
Data was collected using Brandwatch, which includes original tweets, retweets and quote retweets (Ortiz, 2022)

- 26,228 tweets and retweets mentioning the slur **n*gger**, underline triple the 2022 average

- 33,926 mentioning the slur **tr*nny** (against transgender people), up 53% the 2022 average

- 21,903 mentioning the slur **f*ggot** (against gay people), up 39% the 2022 average

- 2,598 mentioning the slur **k*ke** (against Jewish people), up 23% the 2022 average

Elon Musk at the Morgan Stanley Tech, Media and Telecommunications Conference in March 2023:
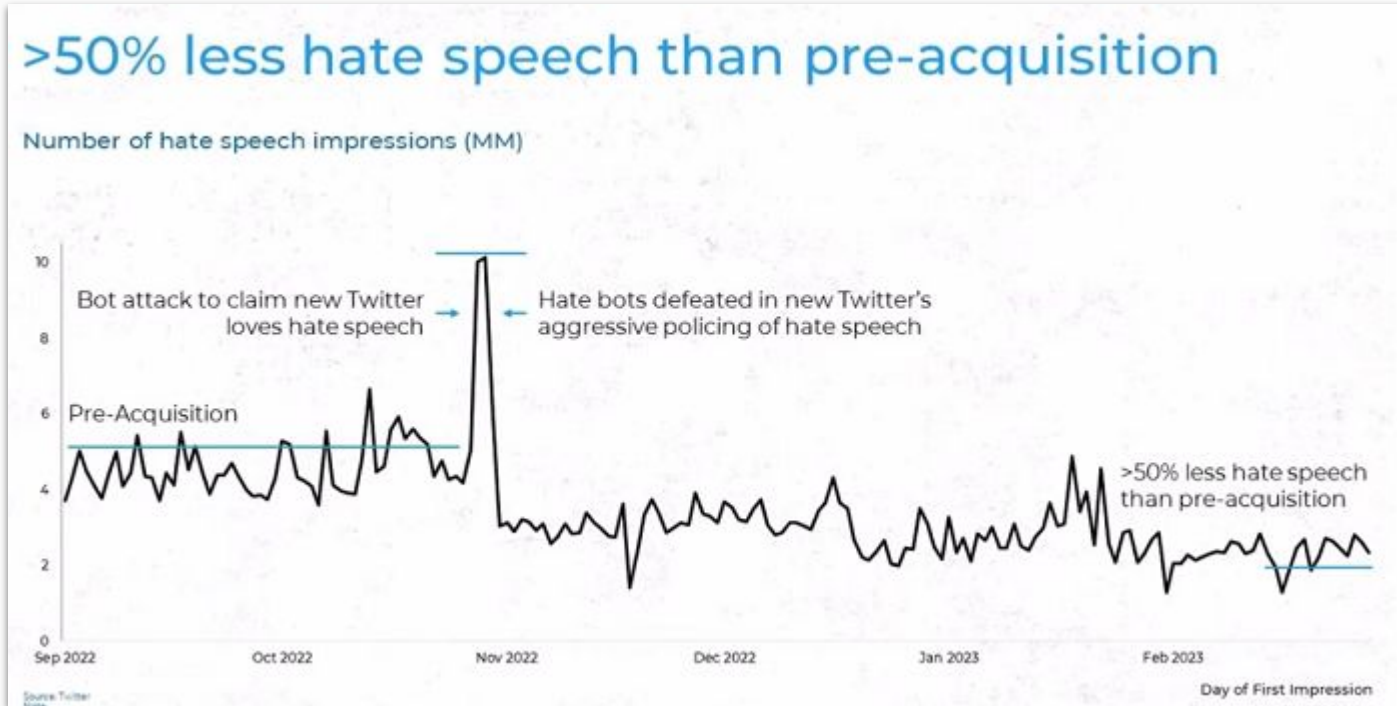
(Hutchinson, 2023)

Research Question

Initial research question: How has hate speech towards the Black community on Twitter evolved since the acquisition by Elon Musk?

Revised research question: Has the amount of hate speech that contains the racist slur "n*gger(s)'' increased after Elon Musk's Twitter takeover?

*Hypothesis: The amount of hate speech that contains the racist slur "n*gger(s)" increased after Elon Musk's Twitter acquisition*

Can this hypothesis be supported by our Twitter data evidence?

Data

Tweets containing the **n word** (c. 10 different ways of writing it)

→ **n*gga** as the most predominantly used, but often in a friendly way among the Black community

→ **n*gger** as mostly pejorative

We chose to focus only on Tweets containing the word "**n*gger(s)**", to **quantify hate speech more accurately**

- Time span for the word count: **10/23/2022** to **11/04/2022**
- Time span for the text classification model: **10/23/2022** to **10/30/2022**

Analysis steps - methodology

- **Data collection** with Twitter explorer collector
- **Pre-process** the text data by removing stop words, punctuation, unnecessary symbols and URLs
- **Tokenization**
- **Sentimental analysis** to see overview of the tweets
- **BERTopic** using non-negative matrix factorization: to look at what topics of hate against black people
- Train the hate speech prediction with Support Vector classifier
- Train the hate speech prediction with **BERT**
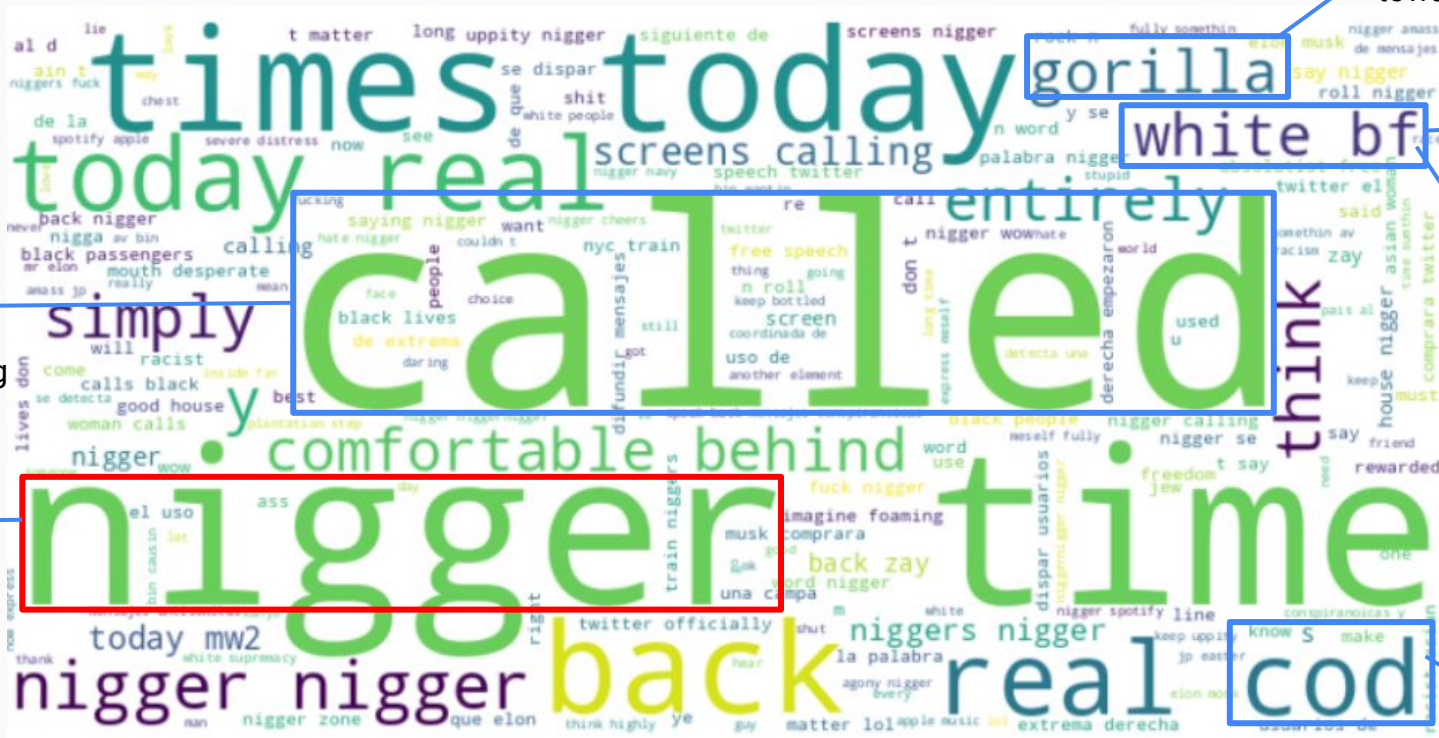
Word count

Tweet counts for nigger OR niggers
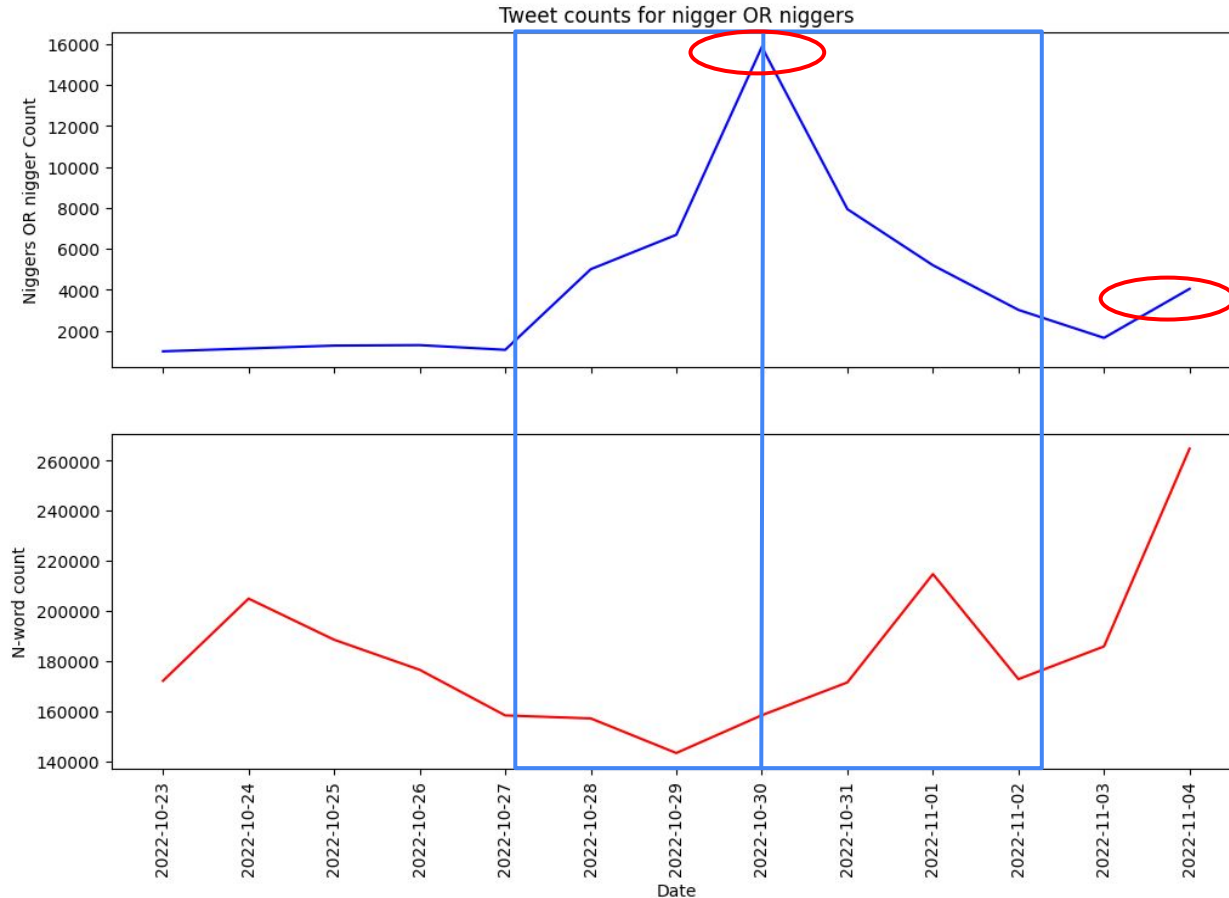
Increase of daily frequency of word "n*gger(s)" in tweets post-acquisition
Spike on Oct. 30th
Increase again on Nov. 4th

All derivations of the n-word, except "nigger(s)"
Used as a proxy of overall number of tweets

Word Cloud

BERTopic
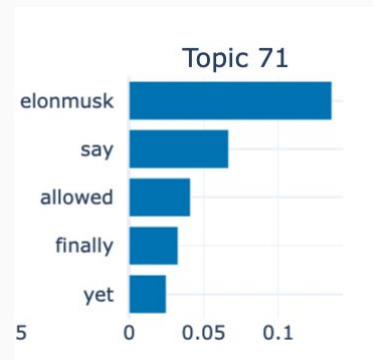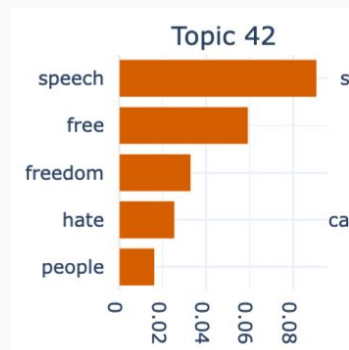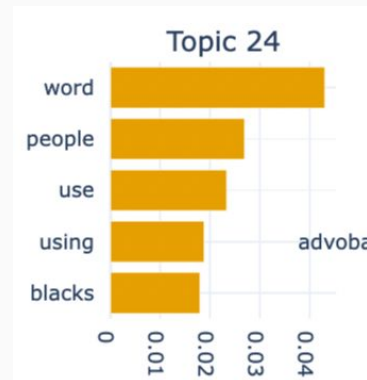
In Total: **99 topics**

Most of them are random and don't give any tangible information. However, the two biggest themes seem to be hate related and to be connected with the decision of Elon Musk to acquire Twitter

Sentiment analysis

Train Sentiment Distribution

Sentiment Analysis of Tweets

Surprisingly, the majority of tweets containing the slur "n*gger" was used **positively**.
Only **11%** of had **negative** sentiment!

Training the model - annotation

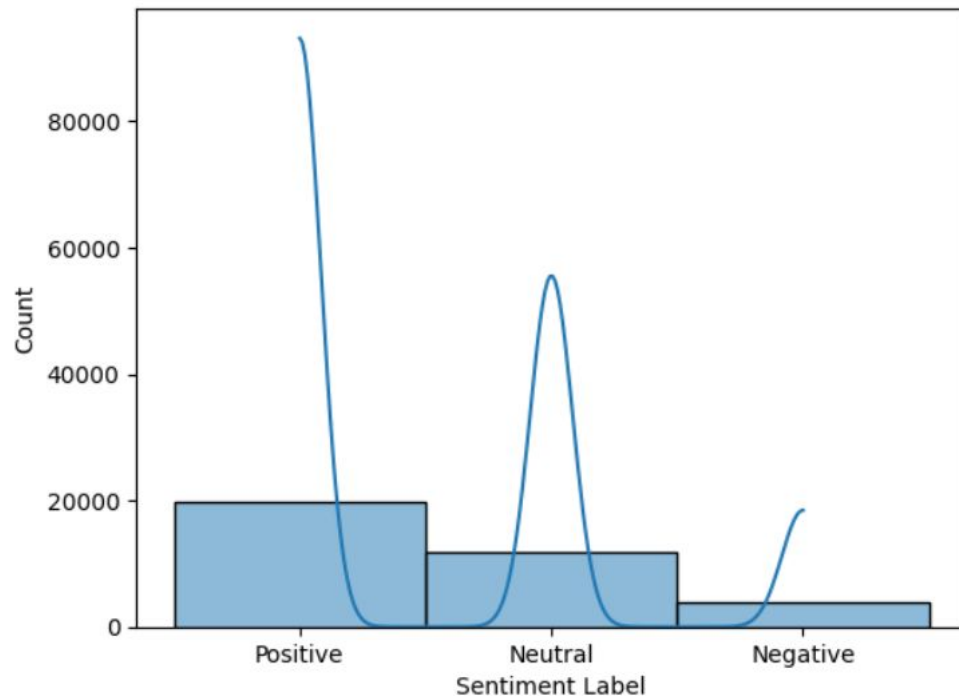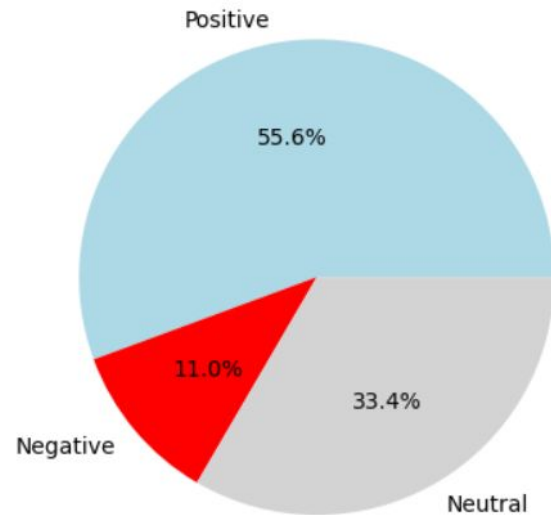| Text	Label
@Ron13Q @PastorDScott No. Nigga is just a lazy way of saying nigger. Black people call each other that racist epitaph because they have diluted themselves into believing it means something different when they say it except when they use the same word as an insult. No difference.	0
@JaMorant F--K YOU NIGGER BLACK SLAVERY	1
Now turn my day into this and my nights no rest crazy ISLAM sand niggers,ya suck like dog ticks with mutt leash's that's like scum leech's for eternity,rotten lazy trash,sit on your asses never work only your women do,being mutt suckers now and forever for getting in my way.	1
@nobuniqqa @Perryinwhite @DriaAlexander @w4bg0rl @562chriss @TuuShiesty @jjasshole Also Negro isn't a slur, Nigga and Nigger are slurs 😂 . Negro is a word.	0
@Untouchable415 @sadedposts what this nigger said	1
"Even when a Native Indian called my father a nigger, I still sympathise with their suffering. Because to identify with my father's formation [as Black], is to be part of no formation at all."—FBWIII	0
trump going defcon 3 on niggers #wagmi #RedWaveComing #MAGA2024	1
@Gregoire1960 @JavonAPrice Johnson called Kennedy's bills "the nigger bills." Kennedy stole his ideas from Eisenhower, who in 1957, at the insistence of Richard Nixon, signed the Civil Rights of 1957 into law. Without that, Kennedy's bills (which expanded on Eisenhower's) would have never happened. -VJ	0
@leoop08 MOTHER FUCKER NIGGER	1
But the nigger in me still kinda wants to try it	0
Nigger is the word today boys	1
@1JovanJordan @HaleySpadesXXX @HaleySpades @1JovanJordanXXX @HushHushEn1 Uh oh! I didn't pay to see that nigger's fat cock inside you 🙈 🙈 🙈 🙈	1
@Naija_PR First of all, we don't talk like that and nigger ain't funny.	0
@kyscuvur Niggers @jisre22	1

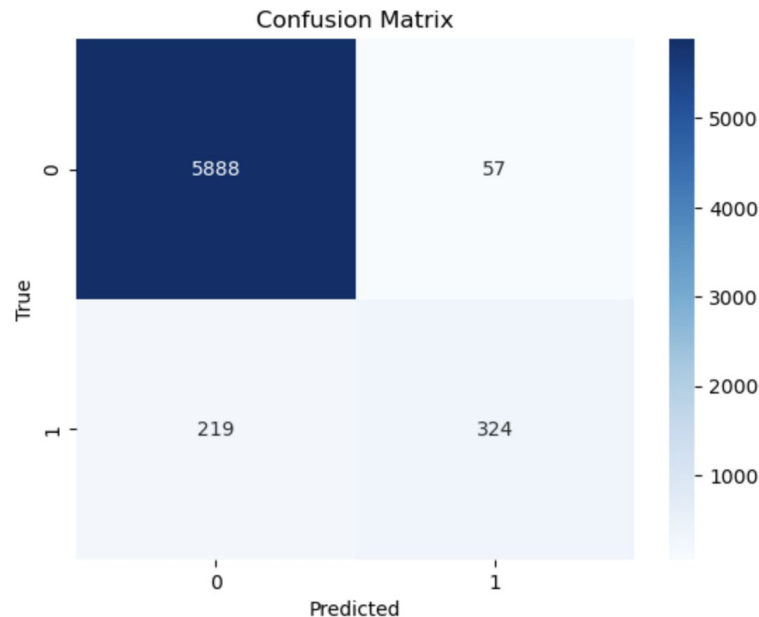| Text	Label
The worst @Airbnb experience just happened to me. The host made an accusation of a party, which I have provided proof of no party last night to both them and Airbnb. The host then decides to refer to me as a nigger and threatens to burn the home down with me in it.	0
OBAMA IM COMING FOR YOU NIGGER	1
All niggers should be banned from Airbnb	1
BIN LADEN NIGGERS « nicolewboyce: Our Taylor Swift Presale War Room Zoom rn https://twitter.com/nicolewboyce/status/1592537205909901313/photo/1 — https://twitter.com/nicolewboyce/status/1592537205909901313 »	1

Hate Speech prediction and its probability

Local Time	Text	Is hateful	Probability
2022-11-01T23:58:26	@lindyli This is a nigger tier take	TRUE	99.39688444
2022-11-01T23:57:34	Damn nigger 😭🤍 « Ambitiouz_Ent: ❤️❤️❤️🥺 #RIPTakeoff https://twitter.com/Ambitiouz_Ent/status/1587482251687854082/photo/1 — https://twitter.com/ambitiouz_er	TRUE	85.23216844
2022-11-01T16:30:37	❤️❤️❤️🥺 #RIPTakeoff https://twitter.com/Ambitiouz_Ent/status/1587482251687854082/photo/1	FALSE	99.93416667
2022-11-01T23:56:18	@GEMINIRIGHTS nigger	TRUE	99.46568012
2022-11-01T23:53:54	RT @FINEASSFATTY: Lmfaoooooo anything to get the niggers back! https://twitter.com/SaycheeseDGTL/status/1587486445786304514	TRUE	81.51456118
2022-11-01T22:36:50	Lmfaoooooo anything to get the niggers back! https://twitter.com/SaycheeseDGTL/status/1587486445786304514	TRUE	96.9463408
2022-11-01T23:53:48	RT @lovedoveclarke: calling me a gorilla and a nigger for simply having a white bf is not the "own" you think it is. y'all are entirely too comfortable behind these screens https://	FALSE	98.29137921
2022-10-31T00:25:11	calling me a gorilla and a nigger for simply having a white bf is not the "own" you think it is. y'all are entirely too comfortable behind these screens https://twitter.com/lovedov	FALSE	96.68251276
2022-11-01T23:52:59	RT @FINEASSFATTY: Lmfaoooooo anything to get the niggers back! https://twitter.com/SaycheeseDGTL/status/1587486445786304514	TRUE	81.51456118
2022-11-01T23:52:35	RT @opinions_hated: @TheSource There is no racial slur that a black person can call a white person that is equivalent to a white person calling a black person "nigger". "Honky"	FALSE	95.03688216
2022-10-31T18:49:50	@TheSource There is no racial slur that a black person can call a white person that is equivalent to a white person calling a black person "nigger". "Honky" and "cracker" don't h	FALSE	60.24754643
2022-11-01T23:52:20	RT @Corrupt_Nigger: The hard fist of justice Corrupt • Police Aggressive • Vulgar Limitless • Account Rt • Mg https://twitter.com/xB:	FALSE	66.03786945
2022-09-01T01:00:44	The hard fist of justice Corrupt • Police Aggressive • Vulgar Limitless • Account Rt • Mg https://twitter.com/xBastard_Flame/status/1	FALSE	99.9687314
2022-11-01T23:51:11	RT @Hushdegod: We got songs saying "I hate niggers" Black men saying that "ima take ya bitch "We are praying chanting our own downfall Simple spells ///and the higher ups are	FALSE	98.9344418
2022-11-01T17:05:58	We got songs saying "I hate niggers" Black men saying that "ima take ya bitch "We are praying chanting our own downfall Simple spells ///and the higher ups are the ones mass fu	FALSE	95.95419168
2022-11-01T23:50:21	RT @monaliza_kc: So if people can say 'nigger' on Twitter these days I definitely can tell someone to "go fuck themselves" right? Testing.	FALSE	99.7600615
2022-11-01T18:59:20	So if people can say 'nigger' on Twitter these days I definitely can tell someone to "go fuck themselves" right? Testing.	FALSE	99.75033402
2022-11-01T23:48:46	RT @CyrusMMcQueen: Twitter in 2022 has resurrected and reimagined the term "White Flight"… You can't be down for the cause only when it's convenient… Black skin = Thick	FALSE	98.54780436
2022-11-01T20:25:34	Twitter in 2022 has resurrected and reimagined the term "White Flight"… You can't be down for the cause only when it's convenient… Black skin = Thick skin… Grow up getting	FALSE	96.48775458
2022-11-01T23:48:19	RT @upblissed: been called a nigger 13 times today. real COD is BACK	FALSE	99.89163876
2022-10-29T23:51:27	been called a nigger 13 times today. real COD is BACK	FALSE	99.83668923
2022-11-01T23:48:08	@ushy101 ive seen this nigger twice on the tl now	TRUE	91.05200768
2022-11-01T23:47:41	RT @irving_twin: My neighbor's four year old stopped me on the street and said, "Elon Musk wants everyone on Twitter to say nigger, and that's a bad word." I replied, "But you j	FALSE	86.81087494
2022-11-01T06:16:01	My neighbor's four year old stopped me on the street and said, "Elon Musk wants everyone on Twitter to say nigger, and that's a bad word." I replied, "But you just said it." Then H	TRUE	55.86785078
2022-11-01T23:47:34	RT @stevozone4_: Reggie Miller broadcasting being a bitch ass nigga about Kyrie Irving smh. He said Kyrie just tweeting a movie is basically the same as two white owners calling	FALSE	97.2194016
2022-11-01T23:43:11	Reggie Miller broadcasting being a bitch ass nigga about Kyrie Irving smh. He said Kyrie just tweeting a movie is basically the same as two white owners calling black people "nigg	FALSE	93.05381775
2022-11-01T23:47:14	Come on don't you want to fucking talk shit you fucking shit nigger	TRUE	99.09473062
2022-11-01T23:46:41	@ordurebizarree Niggers	TRUE	99.54121709
2022-11-01T23:46:32	@lavern_spicer You're a moron, who pals around with MAGAts who call you "nigger," "coon," and "monkey" behind your back.	FALSE	96.25762105
2022-11-01T23:45:54	A man basically called you a black nigger bitch and your response was "not all of us" 😂😂😂	FALSE	86.49704456
2022-11-01T23:45:54	@AnileMassimo Shut up nigger	TRUE	99.52232242
2022-11-01T23:45:46	RT @CyrusMMcQueen: Twitter in 2022 has resurrected and reimagined the term "White Flight"… You can't be down for the cause only when it's convenient… Black skin = Thick	FALSE	98.54780436

Model accuracy and precision with Support Vector Classifier

	precision	recall	f1-score	support
0	0.96	0.99	0.98	5945
1	0.85	0.60	0.70	543
accuracy			0.96	6488
macro avg	0.91	0.79	0.84	6488
weighted avg	0.95	0.96	0.95	6488

Confusion Matrix

Accuracy: 0.9574599260172626

Comparison pre and post-acquisition with SVC

		Hateful	Not hateful
Average number of Tweets per day pre-acquisition (10/24 to 10/26)		**746**	**327**
Average number of Tweets per day post-acquisition (10/28 to 10/30)		**3129**	**4662**
Multiplication factor		**x 4.2**	**x 14.3**

Model accuracy and precision with BERT

```
Average test loss: 0.10
Validation took: 0:00:43
              precision    recall  f1-score   support

           0       0.97      0.99      0.98      5963
           1       0.90      0.65      0.76       525

    accuracy                           0.97      6488
   macro avg       0.94      0.82      0.87      6488
weighted avg       0.96      0.97      0.96      6488
```
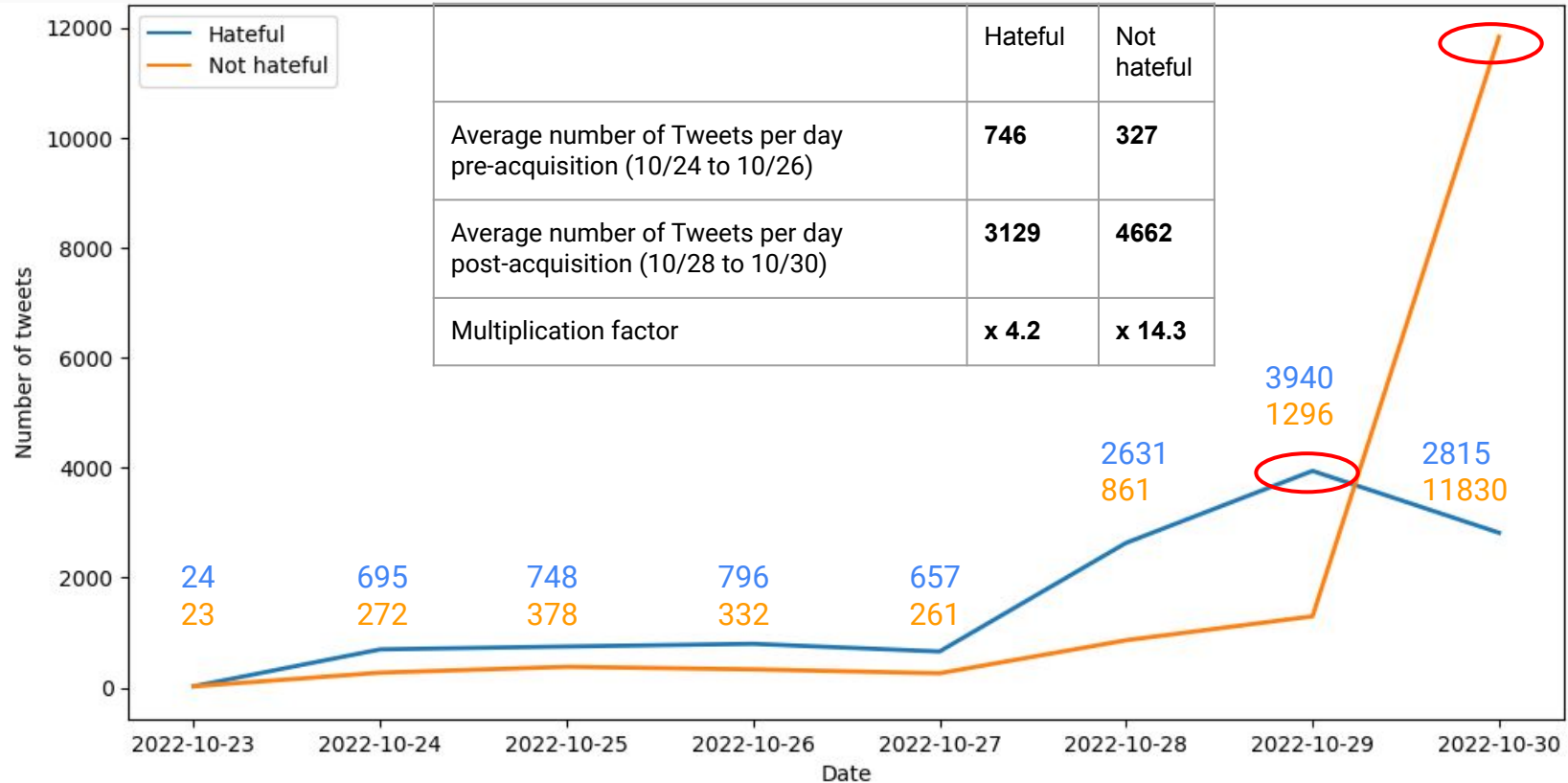
Epoch 1

```
Average test loss: 0.10
Validation took: 0:00:43
              precision    recall  f1-score   support

           0       0.98      0.98      0.98      5963
           1       0.79      0.81      0.80       525

    accuracy                           0.97      6488
   macro avg       0.89      0.90      0.89      6488
weighted avg       0.97      0.97      0.97      6488
```

Epoch 2

```
Average test loss: 0.15
Validation took: 0:00:43
              precision    recall  f1-score   support

           0       0.98      0.99      0.99      5963
           1       0.90      0.75      0.82       525

    accuracy                           0.97      6488
   macro avg       0.94      0.87      0.90      6488
weighted avg       0.97      0.97      0.97      6488
```

Epoch 3

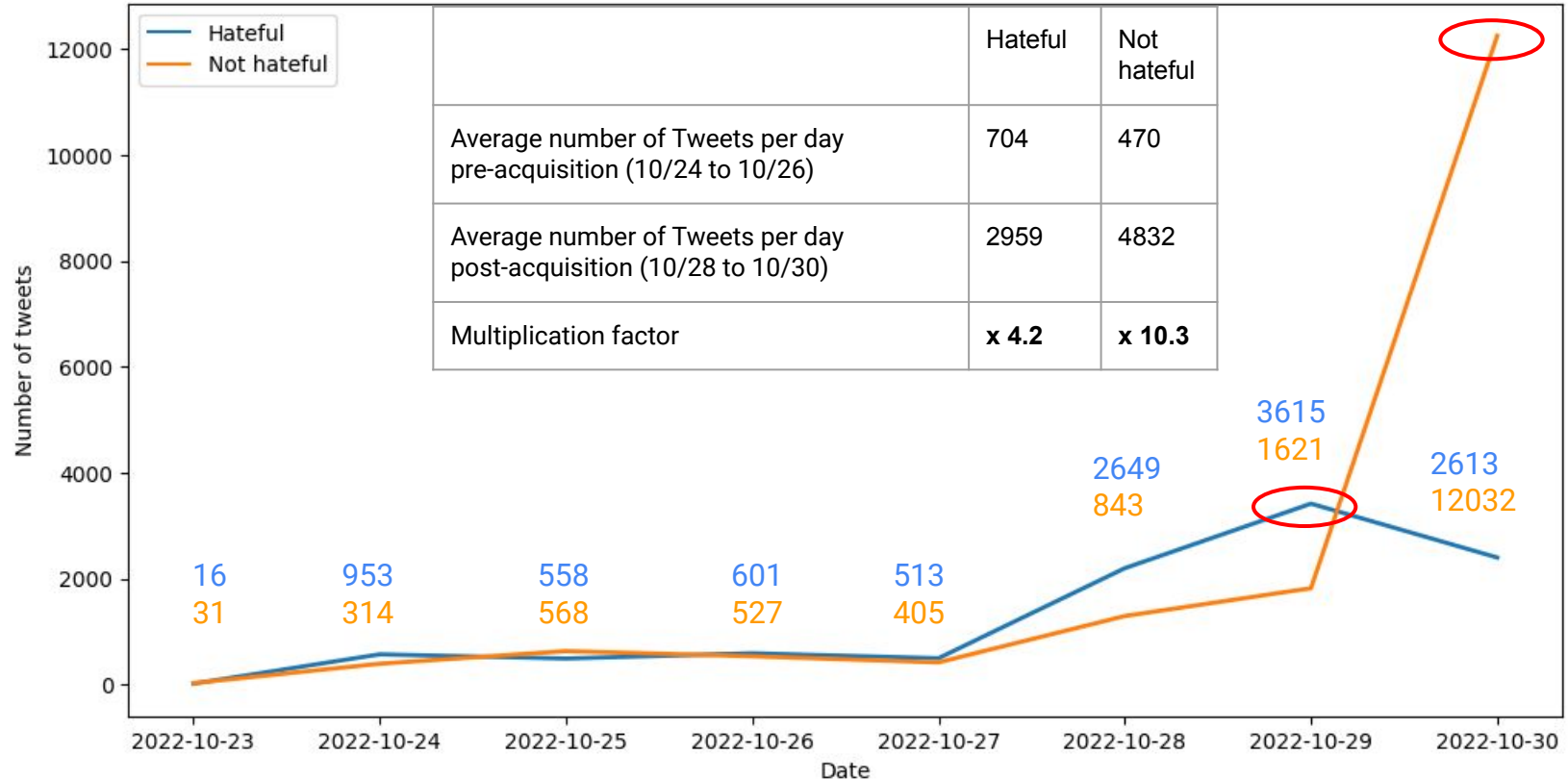```
Average test loss: 0.16
Validation took: 0:00:43
              precision   recall f1-score   support

           0   0.98   0.99   0.99   5963
           1   0.89   0.76   0.82   525

    accuracy   0.97   6488
   macro avg   0.93   0.88   0.90   6488
weighted avg   0.97   0.97   0.97   6488
```

Epoch 4

# Comparison pre and post-acquisition with BERT



|  | Hateful | Not hateful |
|---|---|---|
| Average number of Tweets per day pre-acquisition (10/24 to 10/26) | 704 | 470 |
| Average number of Tweets per day post-acquisition (10/28 to 10/30) | 2959 | 4832 |
| Multiplication factor | **x 4.2** | **x 10.3** |

# Conclusion

**Our research question:** Has the amount of hate speech that contains the racist slur "n*gger(s)" increased after Elon Musk's Twitter takeover?

**Our hypothesis:** Yes, it increased after the acquisition

**Our results:**
- Based on the word count, yes, the frequency of use has increased right after the acquisition and spiked 3 days later, and gradually decreased but staying higher than pre-acquisition.
- Based on our trained model and the SVC model, use in a hateful way increased moderately and spiked 2 days later. On the other hand, use in a non-hateful way increased at a lesser extent during those 2 days, but experienced a surge on October 30th.
- The overall frequency of use of "n*gger(s)" indeed increased on Twitter. However, mere word counting is not sufficient: it is crucial to understand in which context the n-word is being used. We cannot conclude that there is a massive increase in the n-word being used in a hateful way after the acquisition.

# Limitations

- Only the use of "n*gger" was investigated: other n-words can be used in hate speech, and slurs other than the n-word exist too
- We were not able to retrieve the total amount of tweets posted every day
- Annotation for English tweets only, but we were not able to do a language filter through Twitter explorer collector
- Difficult to differentiate between hate speech and non-hate speech in the annotation process without context, which led to many blanks and possible interpretation errors
- Had to reduce the timespan to only 8 days: our initial plan was 2 months (09/27 to 11/27), but the file size was too big
- Google Colab and BERT model were time-consuming and lowered our efficiency
- GPU: limited free use

# Bibliography

Benton, B., Choi, J., Luo, Y., and Green, K. (2022). *Hate Speech Spikes on Twitter After Elon Musk Acquires the Platform*. School of Communication and Media Scholarship and Creative Works. 33.
Available at:
https://digitalcommons.montclair.edu/scom-facpubs/33?utm_source=digitalcommons.montclair.edu%2Fscom-facpubs%2F33&utm_medium=PDF&utm_campaign=PDFCoverPages

Hutchinson, A. (2023). *Elon Musk Outlines His Vision for Twitter at Morgan Stanley Tech Conference*. [online] Social Media Today. Available at:
https://www.socialmediatoday.com/news/Musk-Outlines-Vision-for-Twitter-at-Morgan-Stanley-Conference/644417/.

Ortiz, A.G. (2022). *Musk's claim about a fall in hate speech doesn't stand up to scrutiny*. [online] Center for Countering Digital Hate | CCDH. Available at:
https://counterhate.com/blog/fact-check-musks-claim-about-a-fall-in-hate-speech-doesnt-stand-up-to-scrutiny/