Using Machine Learning To Identify Threat Actors Online

A look at extremism, machine learning, and research approaches

Agenda

Identifying and Understanding Violent Extremism on Parler

1. Understanding **Threat Actors** outside of the typical criminal, hacktivist, state actor vectors.
2. Applying a **Research Mindset** to cyber security challenges, problems, and day-to-day operations.
3. Using **Machine Learning** for classification, prediction, and processing of large datasets.

1. This talk comes with a **Content Warning** for violent far-right extremism. One sample of direct text content is shared.
2. The research in this talk focuses on **far-right extremism**, however, the research can be applied to far-left and other forms of ideological extremism.

James Stevenson (He/ Him)

JamesStevenson.me

Background:

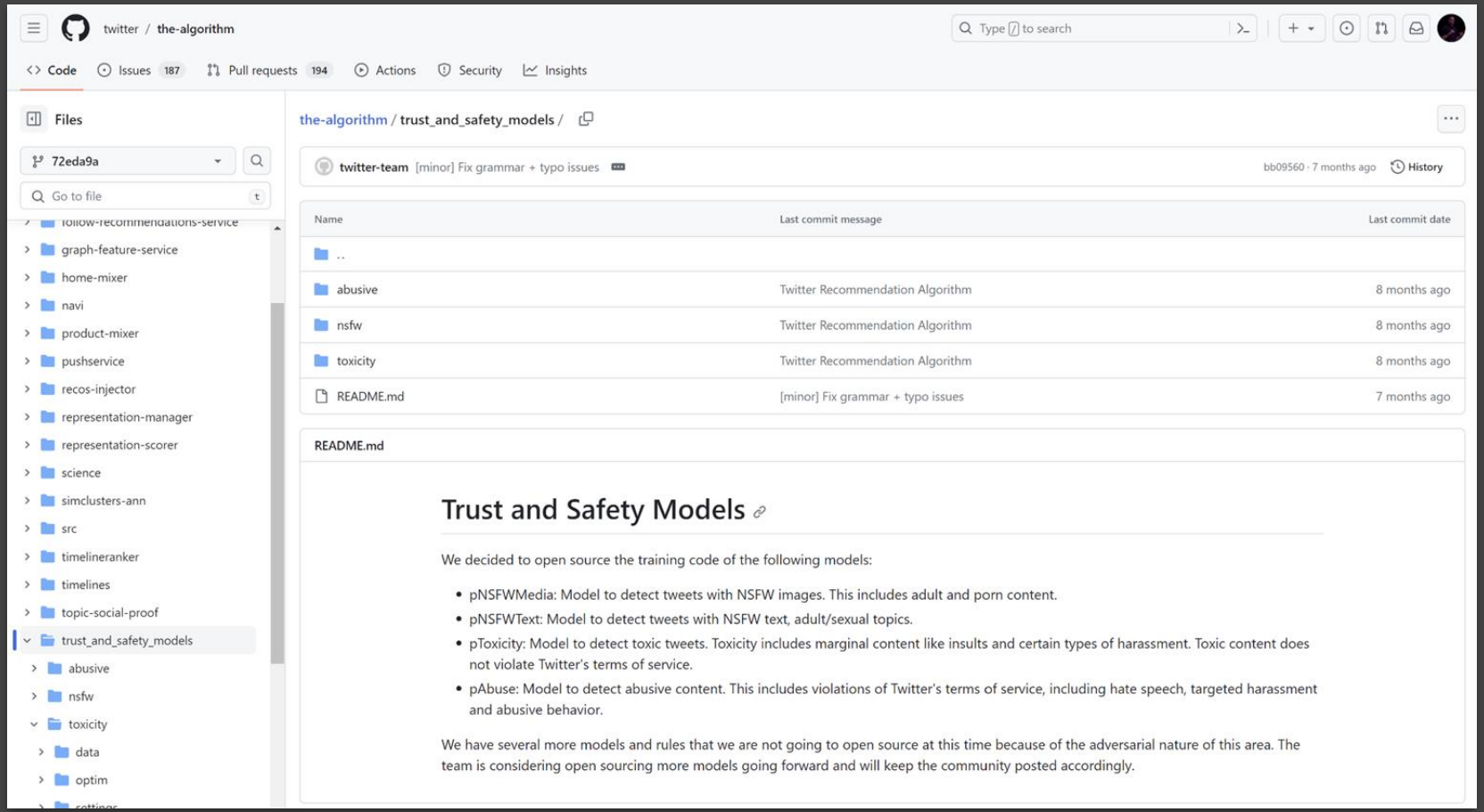
- **SOC Analyst**
- **Android Internals Software Engineer**
- **Offensive Security Research**
- **Vulnerability Research**

Side Projects:

- **Published and Self-Published Author**
- **Part-time PhD Student**
- **Occasional Conference Speaker**

Facebook | Statista

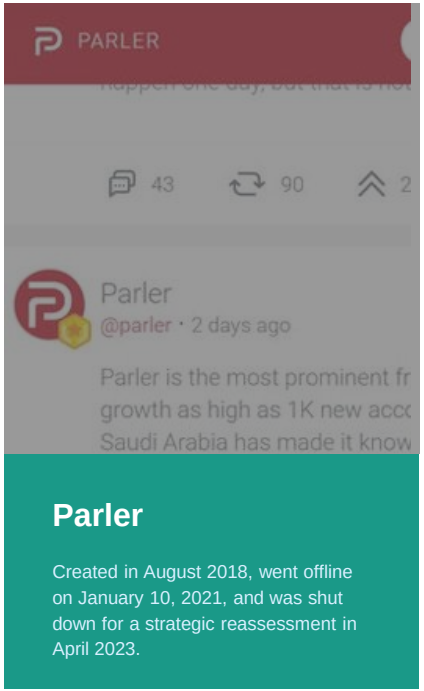
In the second quarter of 2023, Facebook acted on 13.6 million pieces of terrorism content, down from 14.5 million pieces of content actioned in the first quarter of 2023

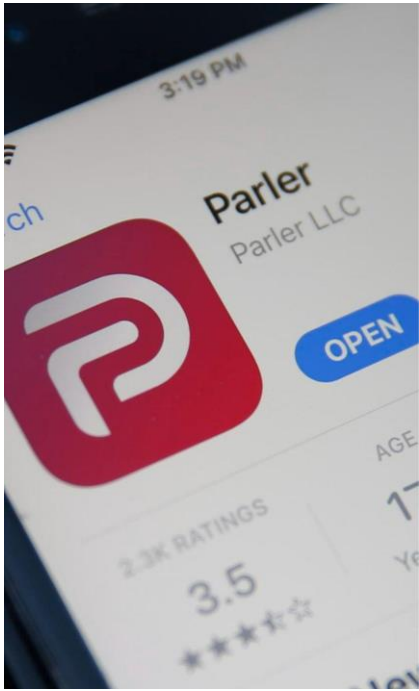


https://github.com/twitter/the-algorithm/tree/72eda9a24f815f6d566818cbf8518138e29d83e9/trust_and_safety_models

Social Networks
Parler and Stormfront

Datasets:
5000 unprocessed forum posts originating from the Stormfront*1 platform posted between 2002 and 2017, and **183M Parler posts** made by 4M users posted between August 2018 and January 202*2.





*1 https://github.com/Vicomtech/hate-speech-dataset

*2 https://arxiv.org/pdf/1905.08067.pdf

Terrorism and the internet: How dangerous is online radicalization?

“Online radicalization can and does occur, with potentially violent consequences”

Global Terrorism Index| Institute for Economics and Peace (IEP)

In 2019 there was a reported rise of 320% in the total number of far-right terrorism incidents in the West - particularly in Western Europe, North America, and Oceania

Online-Enabled Violent Far-Right Extremism

We define **online-enabled far-right extremism** as “the use of social media to spread, incite, demonstrate, or plan activities motivated by an individual's extreme right-wing beliefs”.

The use of social media

Spread, incite, demonstrate, or plan violent activities

Motivated by an individual's extreme right-wing beliefs

Research Approach

Make Observations

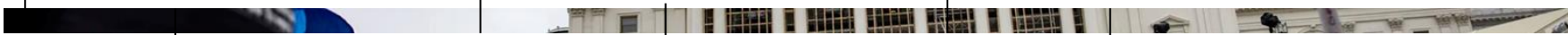
Comprise An Interesting Question

Formulate Hypothesis

Produce Testable Predictions

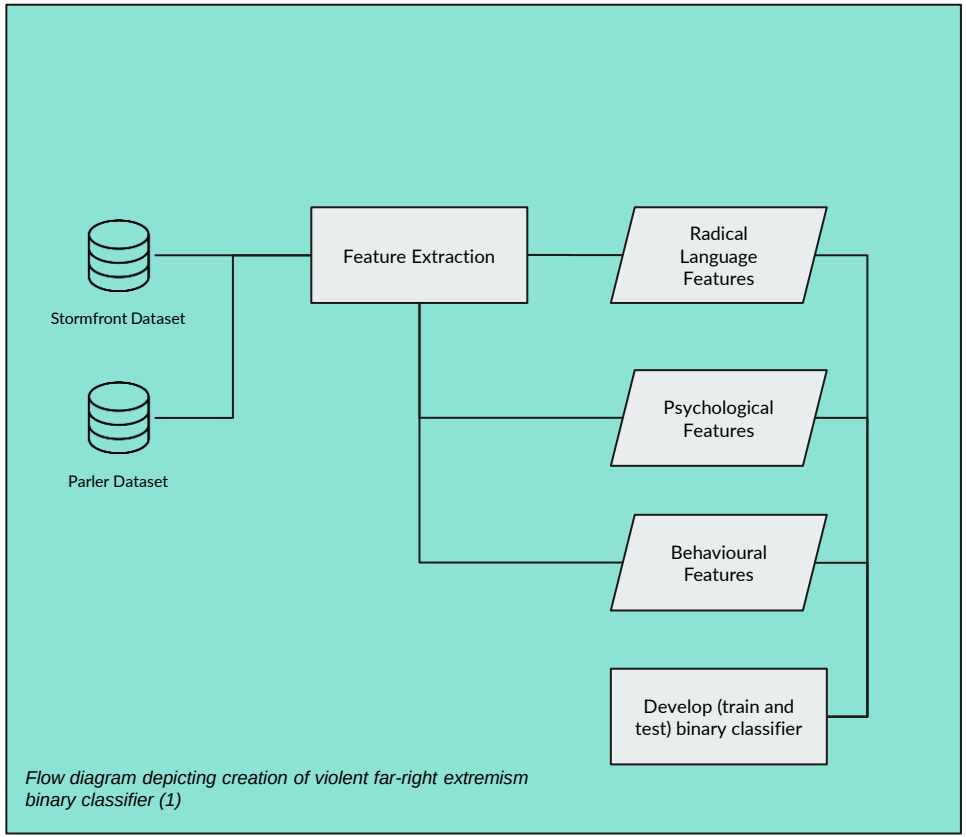
Gather Data To Test Predictions

Develop General Theories

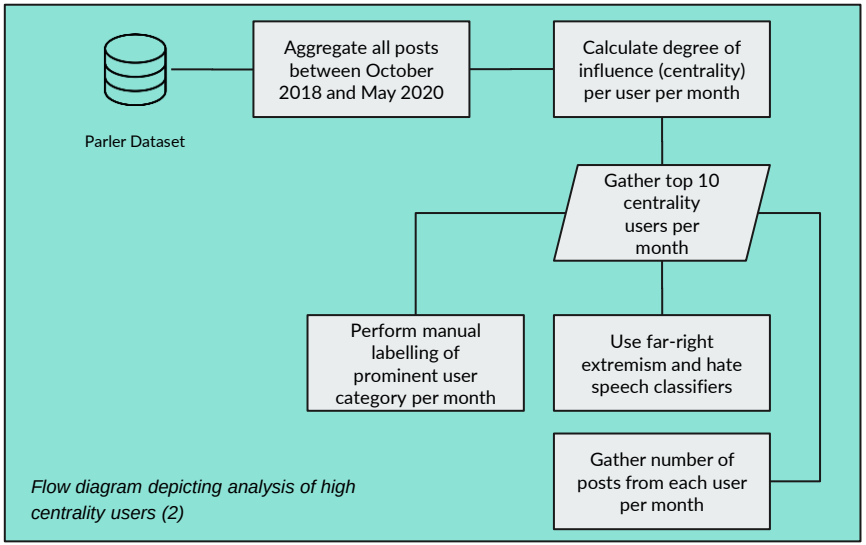


Research Questions and Methods

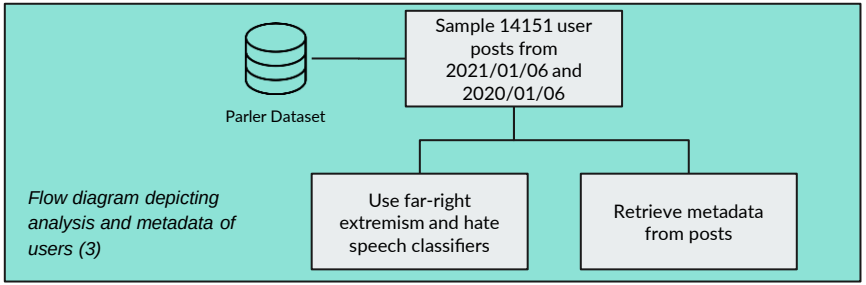
Can methods used to identify Islamic extremism text posts be applied to far-right extremism?	Build a classifier to identify far-right extremism. (1)
What are common activities performed by users on the Parler social network?	Manually identifying distinctive user behaviours visible within Parler. *(2)*
Is the Parler social network primarily used to share hate speech and far-right extremism?	Analyse prevalence of far-right extremism in posts from Parler users and overall levels of engagement with extremist content. (3)



Flow diagram depicting creation of violent far-right extremism binary classifier (1)



Flow diagram depicting analysis of high centrality users (2)



Flow diagram depicting analysis and metadata of users (3)

Approaches

Radical Language Features

Classifier Features

We establish a baseline for how far-right extremist messages are constructed. Both by using the Stormfront dataset and secondly with violent / hate speech dictionary lists.

- Word Vector Embedding scores
- Capital Letter and Violent word frequency

Behavioural Features

Classifier Features

Individuals leave indicators of their personality online based on their day-to-day choices, including word choice (LIWC scores) . We use this word choice as a feature set.

- LIWC Dictionary Scores
- Minkowski distance

Psychological Signals

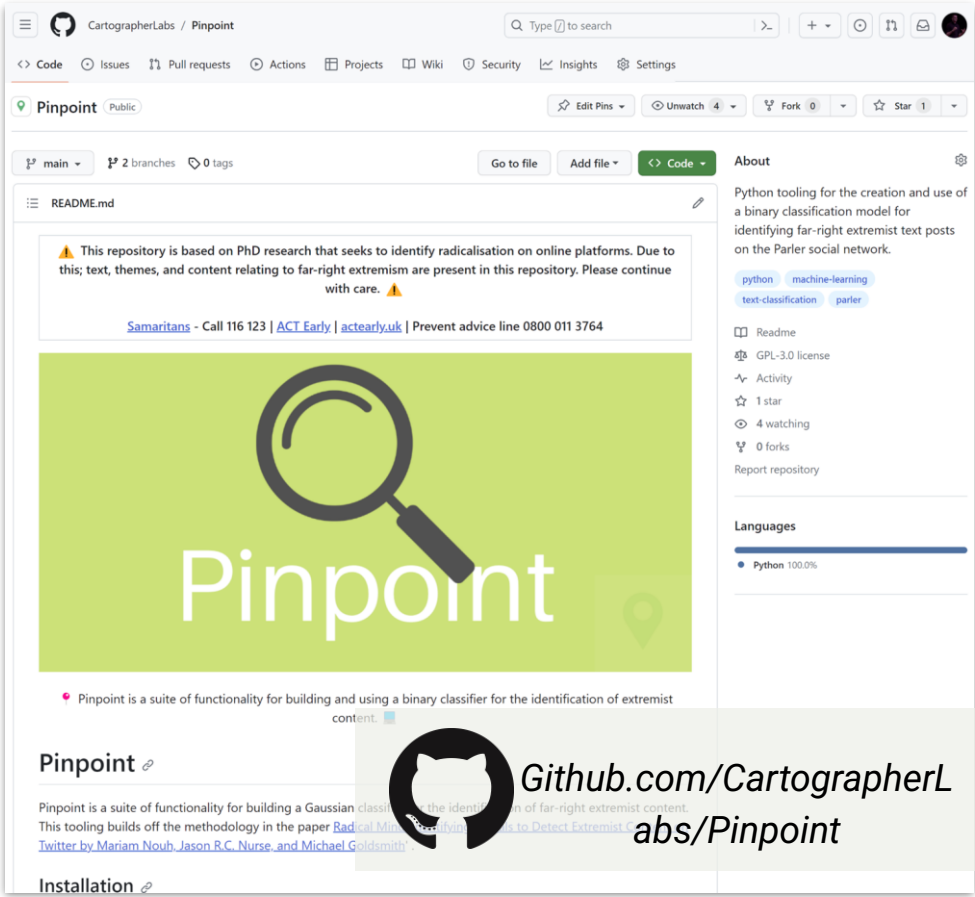
Classifier Features

Behavioural features relate to how a specific individual acts and portray themselves online. To capture how users interact with other users, we used a mention interaction graph.

- Degree Centrality (Mention Graph)

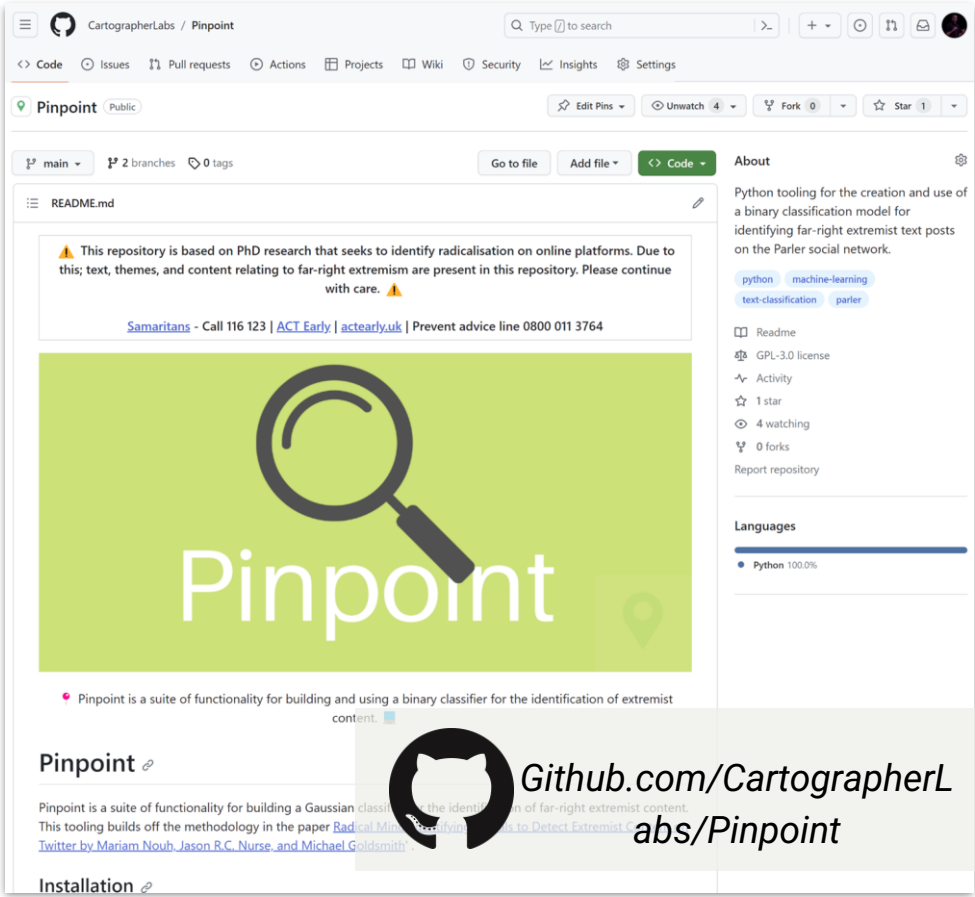
Building A Classifier

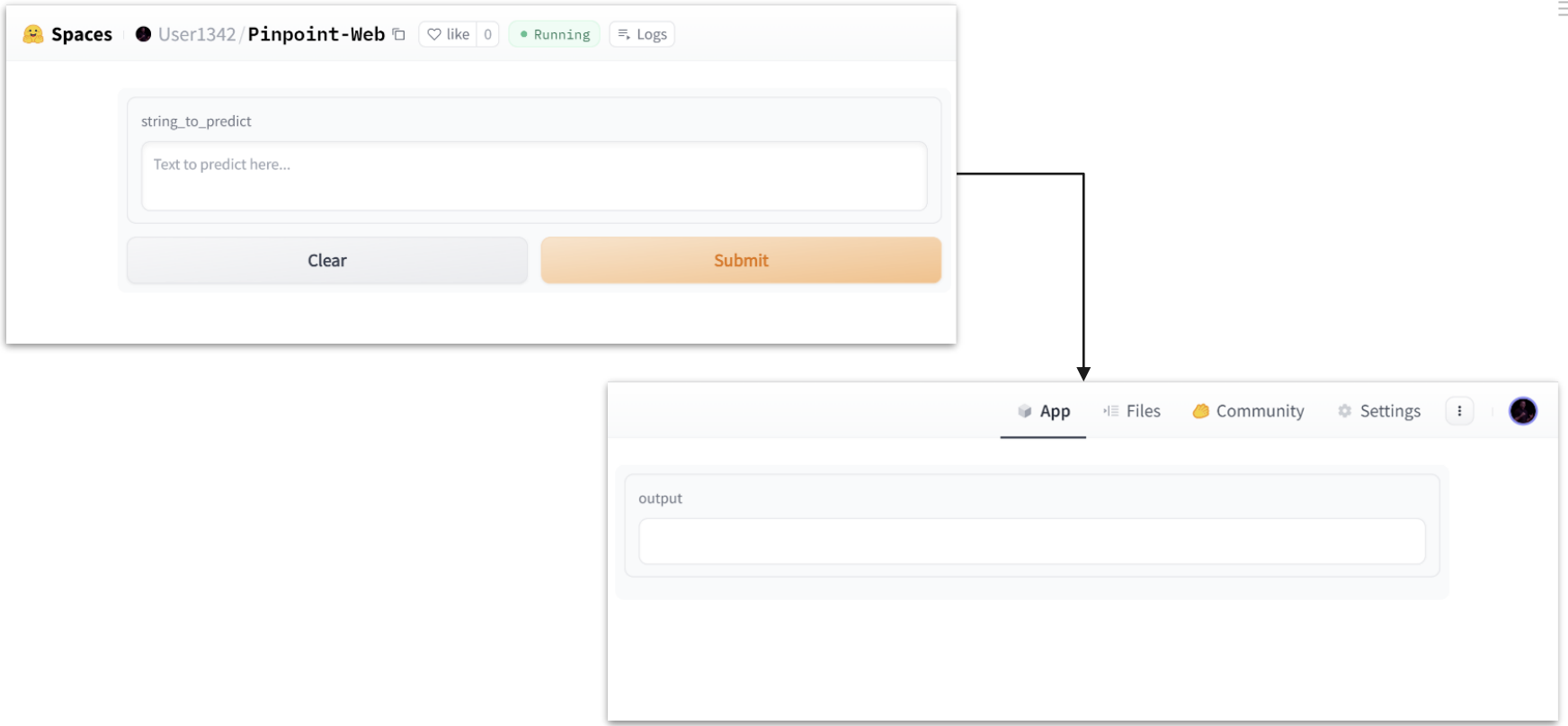
- FeatureExtraction.py
  - Aggregator_NGram.py
  - Aggregator_TfIdf.py
  - Aggregator_Word2Vec.py
  - Aggregator_WordingChoice.
  - Grapher.py
- RandomForest.py
- Sanitizer.py
- Serializer.py



Dependencies

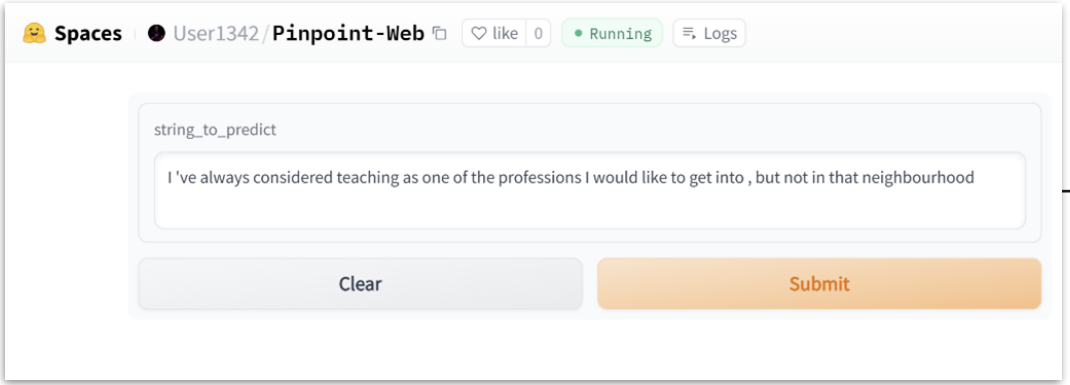
- Gensim
- Scipy
- Pandas
- Numpy
- Networkx

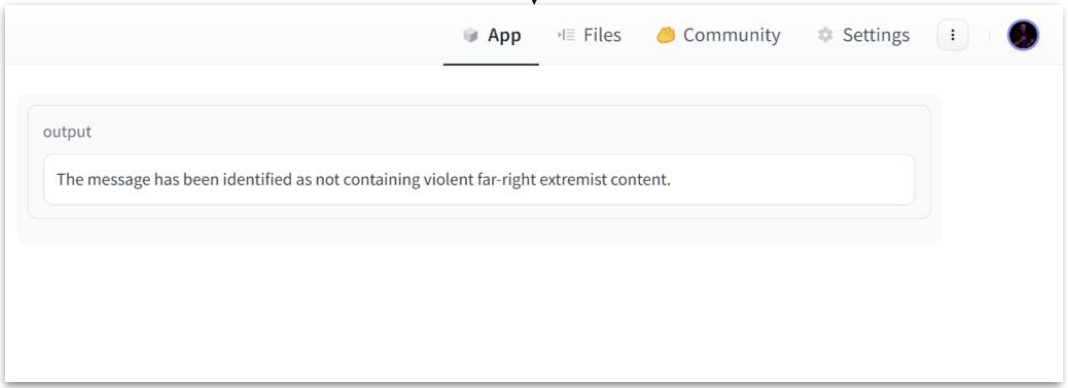




Pinpoint Web Demo (Extremism Classifier)

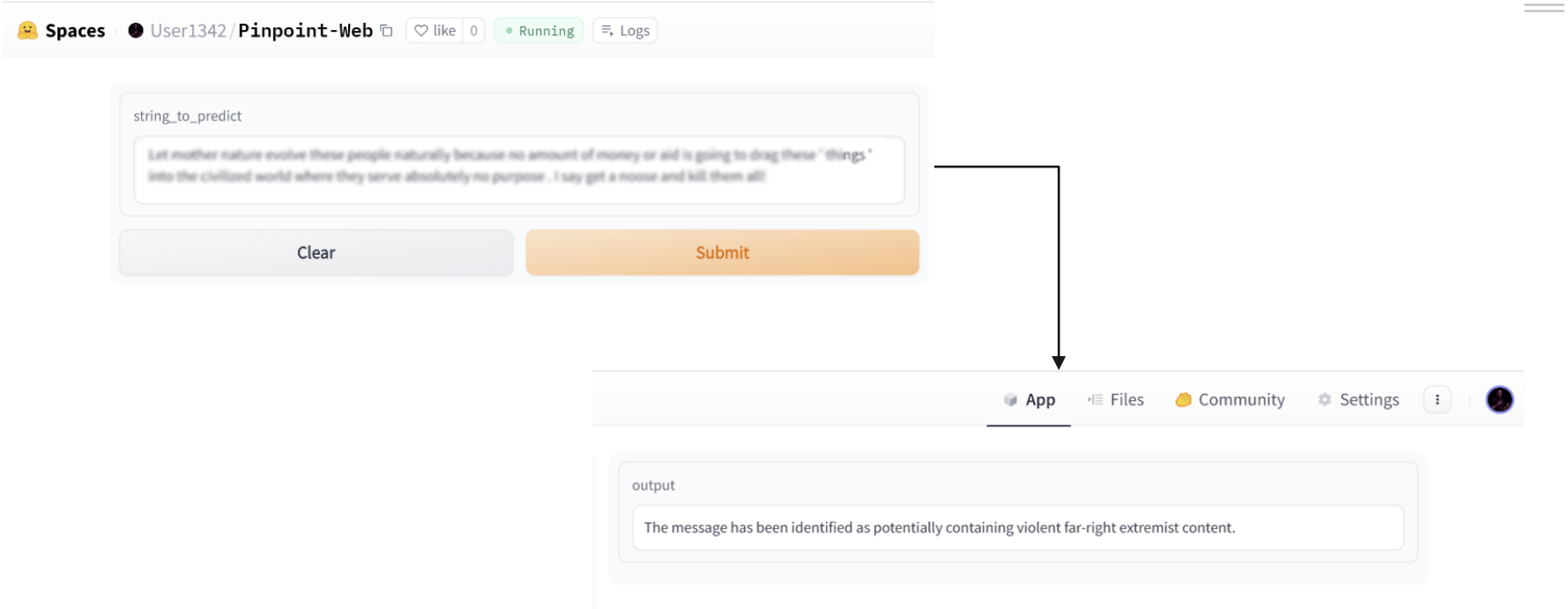
HuggingFace.co/spaces/**User1342/Pinpoint-Web**





Pinpoint Web Demo (Extremism Classifier)

HuggingFace.co/spaces/**User1342/Pinpoint-Web**



Pinpoint Web Demo (Extremism Classifier)

HuggingFace.co/spaces/**User1342/Pinpoint-Web**

Classifier Performance *(1)*

Type	Accuracy	Recall	Precision	F-Measure
All	**0.789**	0.697	0.839	0.762
Behavioural	0.693	1	0.61	0.758
Psychological	0.742	0.691	0.755	0.722
Language	0.736	0.627	0.784	0.697

Type	Feature Weight
…	
violent freq	1.46%
analytic	1.64%
power	1.96%
minkowski	2.69%
anger	3.81%
centrality	4.56%
message vectors	**73.76%**

Conducting

Hate Speech and Inciting Violence

Parler Accounts

Political, News, or Other Commentary

Conspiracy Theorists

News or External Social Media Aggregation

Anti-Typical-Parler User / Anti-Far-Right

Actors *(2)*

7 Prominent Activity Categories

Extremist and Hate Speech Prevalence *(3)*

Extremism In Posts

29.5%

Percentage of sampled posts identified as extremist

Toxicity and Extremism

0.298

Weak positive correlation coefficient between post toxicity and extremism

Average Toxicity*

17.7%

Average toxicity across all sampled posts

*https://github.com/unitaryai/detoxify

What This Tells Us and Why It Matters

Parler was a social network where hate speech and violent **extremist rhetoric thrived** due to the lack of moderation. As well as serving as a platform for sharing right-wing views, connecting like-minded individuals.

Research

29.5%

Percentage of sampled posts identified as extremist

Machine Learning

79%

Far-right extremism classifier accuracy

Threat Actors

7

Identified user groups/ categories for high centrality users

MOTHERBOARD
TECH BY VICE

The Nazi-Free Alternative to Twitter Is Now Home to the Biggest Far Right Social Network

Gab, which has been used frequently by neo-Nazi terror groups to organize and recruit, is now the biggest node on the Mastodon netw

By Ben Makuch

11 July 2019, 8:01pm Share Tweet Snap

The Verge / Tech / Reviews / Science / Entertainment / More +

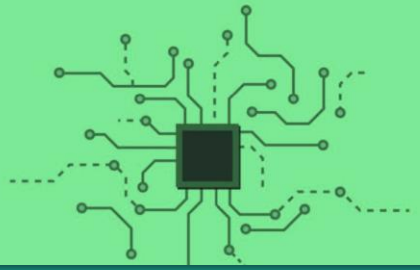
REPORT

How the biggest decentralized social network is dealing with its Nazi problem

Mastodon was built to be a kinder, more decentralized version of Twitter – then Gab showed up

ANALYSIS: THE USE OF OPEN-SOURCE SOFTWARE BY TERRORISTS AND VIOLENT EXTREMISTS

September 18, 2019



How do smaller and indie networks handle extremism?

Closing Thoughts And Wrap Up

1. Understanding **Threat Actors** outside of the typical criminal, hacktivist, state actor vectors.

2. Applying a **Research Mindset** to cyber security challenges, problems, and day-to-day operations.

3. Using **Machine Learning** for classification, prediction, and processing of large datasets.

Github.com/user1342

DRAFT

Analysing The Activities Of Violent Far-Right Extremists On The Parler Social Network

James Stevenson
Matthew Edwards
Awais Rashid

Bristol Cyber Security Group
School of Computer Science
University of Bristol
Bristol, UK
Email: {my19303,matthew.john.edwards,awais.rashid}@bristol.ac.uk

***Abstract*—A significant gap remains in our understanding of the types of users who utilise online extremist platforms, as well as how their activity on these platforms influences the radicalisation of others and the dissemination of extremist content online. Our research addresses this gap by focusing on the Parler social network, one of the largest social media platforms used by the extreme far-right, boasting a reported 15 million total users as of January 2022. We present an exploration of the Parler social network, specifically reviewing the roles of users in the network and the types of activity and content shared on the platform. Our methodology provides a novel examination of Parler using tools that have previously been tested to understand other extremist groups.**

I. Introduction

In 2019, The Global Terrorism Index, published by the Institute for Economics and Peace (IEP), reported a 320% rise in the total number of far-right terrorism incidents in the West, particularly in Western Europe, North America, and Oceania [16]. This threat is not only present within the US, but also in other western nations such as the UK. In 2021, MI5 Director General Ken McCallum stated that while right-wing terrorism was not at the same scale as Islamic terrorism, it was, however, growing, and of the 29 late-stage attack plots between 2018 and 2021, 10 had been extreme right-wing based [20].

With the rise of internet communication, a large proportion of extremist activities now utilises online communities for communication, including sharing extremist content, radicalising others, and planning terrorist actions. In Q1 of 2023 alone, Facebook took action on approximately 14.5 million pieces of terrorist content on their platform [27]. However, due to the ease of creating such radicalised content on these platforms, these extremist communities continue to be popular and grow.

As of January 6th, 2021, Parler had approximately 15 million total users [14], and the platform was also endorsed far-right extremists. Given these circumstances, our method focuses on the Parler social network as a cornerstone and representative sample of extreme far-right social media usage.

While previous research contributions have broadly explored the types of far-right extremist groups and provided a first-look analysis of platforms such as Parler and Gab, this study provides novel contributions and key insights by answering the following research questions:

- Can methods used to identify Islamic extremism text posts be applied to far-right extremism?
- What are common activities performed by users on the Parler social network?
- Is the Parler social network primarily used to share hate speech and violent far-right extremism?

II. Related Work

Much research has already been performed by the social science research community into far-right extremism. This has included the effects of far-right extremism on society [22] [29] [31], the differences between far-right extremist groups [7] [9] [12], and how far-right extremism has risen in popularity [1]. Much work has also been performed by the computer science research community in using data driven approaches to identify terrorism online [10]. This paper builds upon these approaches by both developing a machine learning model for the automated identification of far-right extremist content online, and by exploring the types of activities performed by users on the Parler social network, helping explain the various uses of far-right online platforms.

Previous research has attempted to understand the structure and population of far-right extremist platforms, particularly Gab, Parler, and Stormfront [19], [30]. Aliapoulios et al. [2] present a dataset of 183M Parler posts made by 4M users between August 2018 and January 2021, as well as metadata

Github.com/user1342

JamesStevenson.me/paper

Common Questions

Did this research go through an ethics application?

Why does this research focus on far-right extremism?

Where did you get your datasets from?

Closing Thoughts And Wrap Up

1. Understanding **Threat Actors** outside of the typical criminal, hacktivist, state actor vectors.

2. Applying a **Research Mindset** to cyber security challenges, problems, and day-to-day operations.

3. Using **Machine Learning** for classification, prediction, and processing of large datasets.

Github.com/user1342