



detecting and
preventing
CYBERBULLYING in the
interest of
CHILD SAFETY

CYBERBULLYING DETECTION

BY DIANE
TUNNICLIFFE

Trigger Warning

This presentation contains analysis of offensive and hateful language.
If you find yourself triggered by toxic words and phrases, please take
this opportunity to exit the presentation.

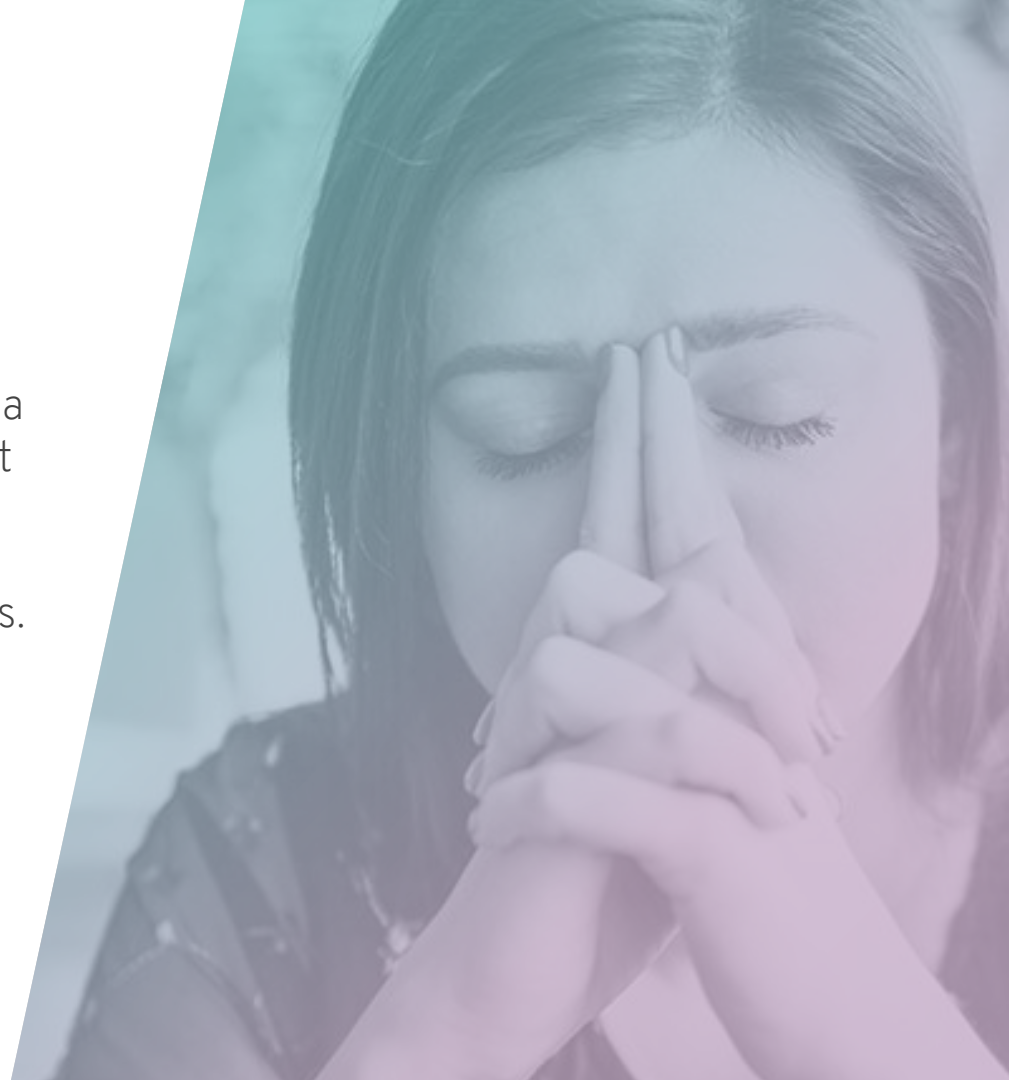
TABLE OF CONTENTS

- 01 PROBLEM STATEMENT
- 02 BUSINESS VALUE
- 03 METHODOLOGY
- 04 RESULTS
- 05 RECOMMENDATIONS
- 06 FUTURE WORK

Problem Statement

Children targeted via cyberbullying report depression, anxiety, suicidal behavior, and other symptoms.

By predicting cyberbullying and stopping it before it happens, we can protect our kids.

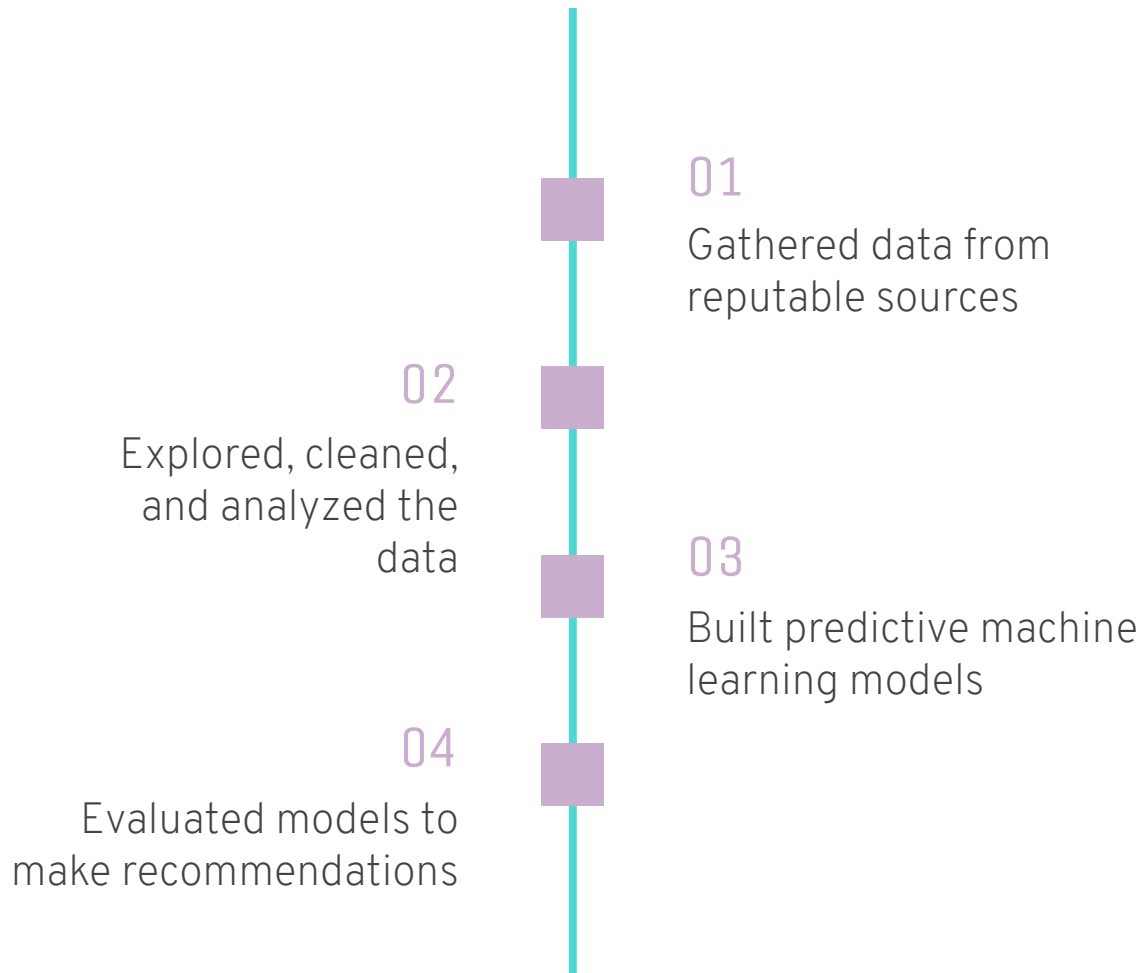


BUSINESS value

- Keep social media platforms safe.
- Protect children and ensure that children can safely use platforms.
- What are the trends in this realm?
- Can a model predict cyberbullying?
- How can we put this to use?
- **Safety => continued use => continued profit!**



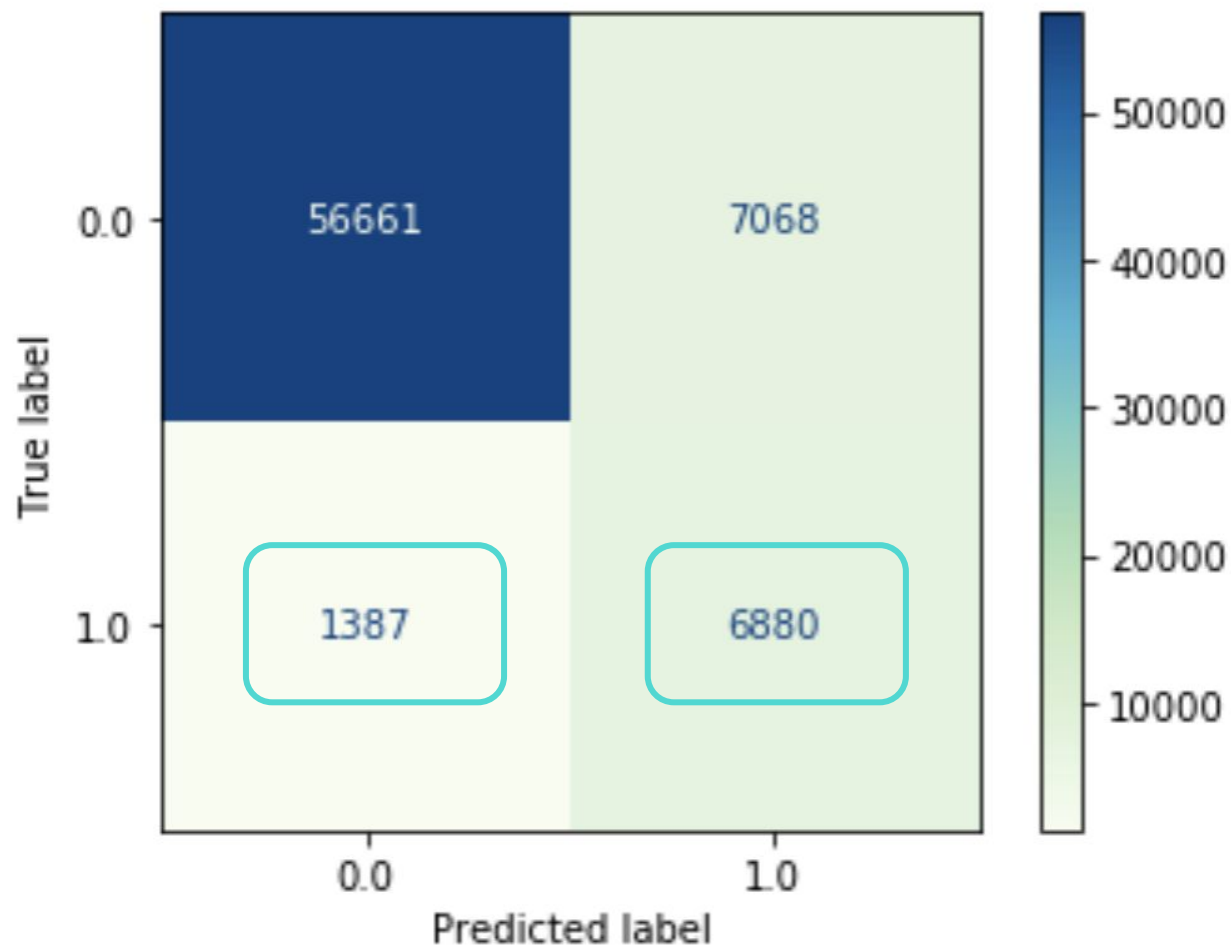
METHODOLOGY



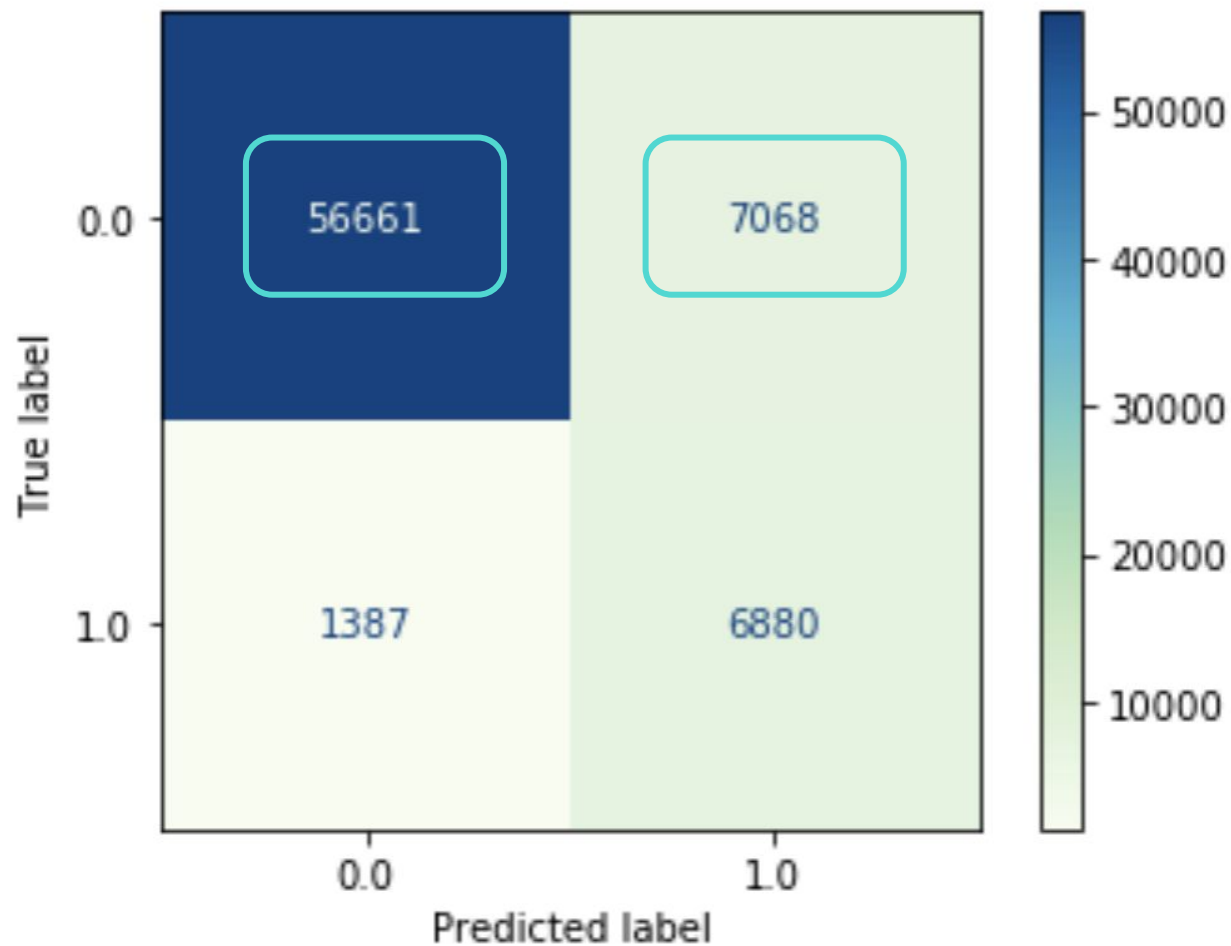
NON-CYBERBULLYING MOST COMMON WORDS



results



results



evaluation

83%

RECALL/
SENSITIVITY

ACCURACY

88%

BASIC example

```
predict(['you are so beautiful'])
```

not cyberbullying

```
predict(['you are so ugly'])
```

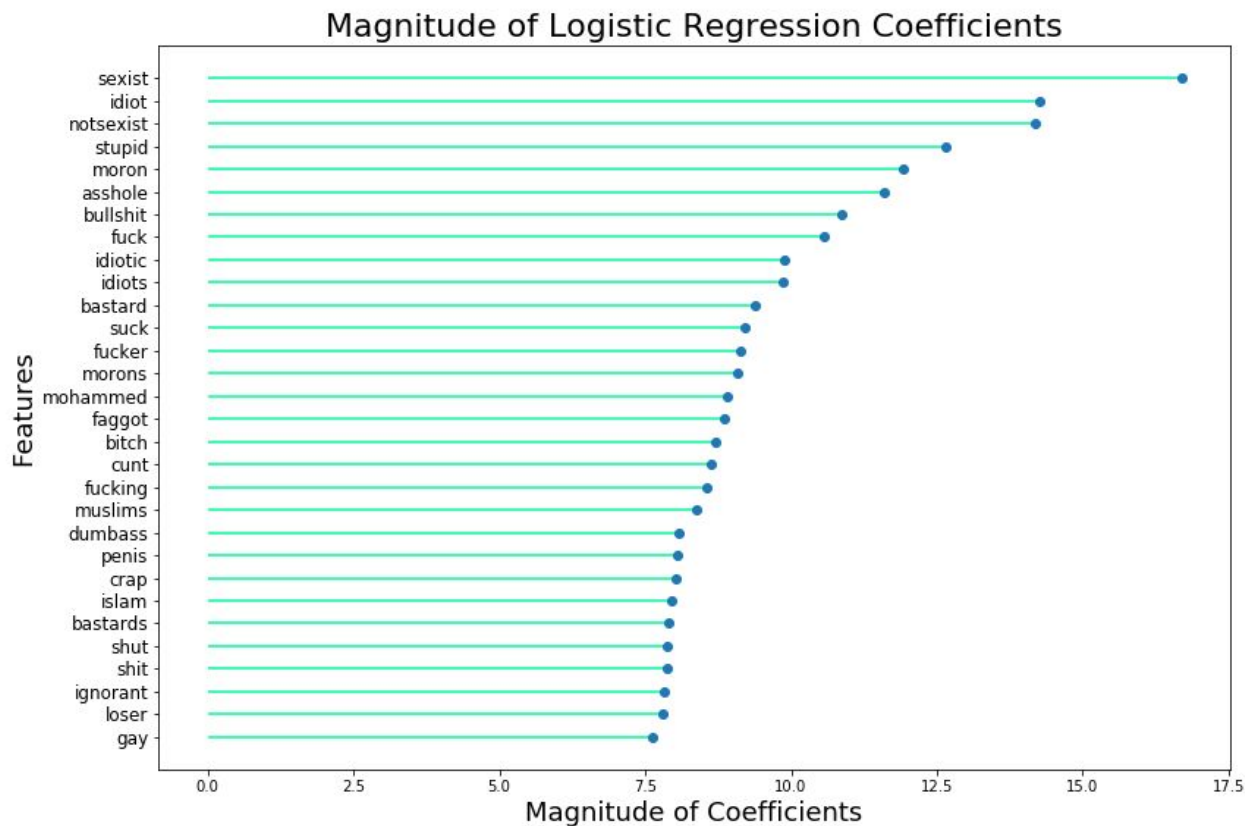
cyberbullying

USE a GOOD MODEL

Fast, efficient,
trained to detect
cyberbullying.



results



BLOCK BAD WORDS

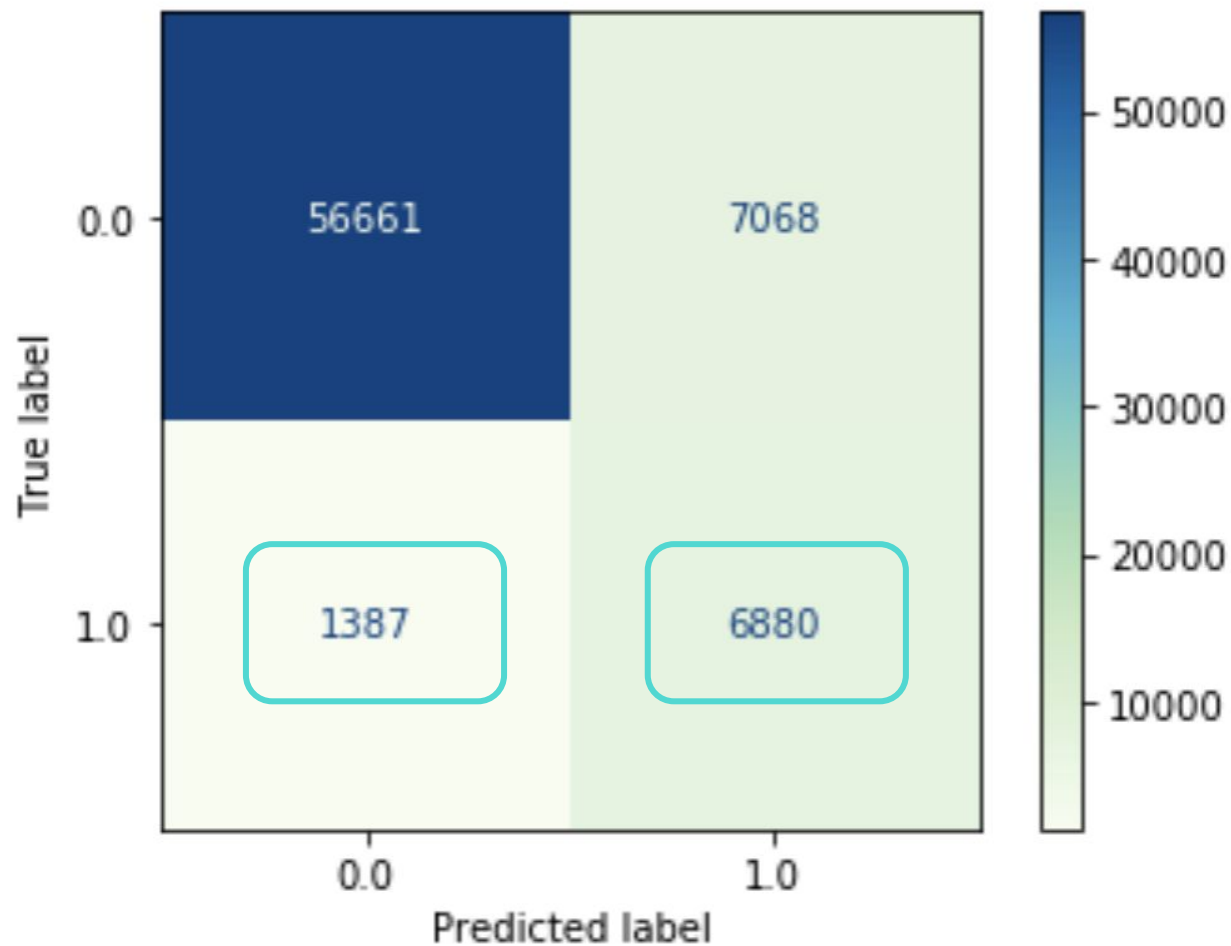
Flag/filter/ block the words
that are strong predictors of
cyberbullying

USE A GOOD MODEL

Fast, efficient,
trained to detect
cyberbullying.



results



BLOCK BAD WORDS

Flag/filter/ block the words that are strong predictors of cyberbullying.

USE a GOOD MODEL

Fast, efficient, trained to detect cyberbullying.

Prevent!

Ensure continued profit by fully focusing on prevention.



FUTURE WORK

01

See if examining word contexts improves the model's ability to predict cyberbullying

EXPLORE CONTEXT

02

The model we have performs great! But can another type of model possibly do better?

NEW MODELS

03

Gather even more data to train the model on.

MORE DATA

THANK YOU!

Do you have any questions?

diane.j.tunncliffe@gmail.com

github.com/dtunncliffe

[https://www.linkedin.com/
in/diane-tunncliffe](https://www.linkedin.com/in/diane-tunncliffe)

CREDITS: This presentation
template was created by
Slidesgo, including icons by
Flaticon, and infographics &
images by **Freepik**

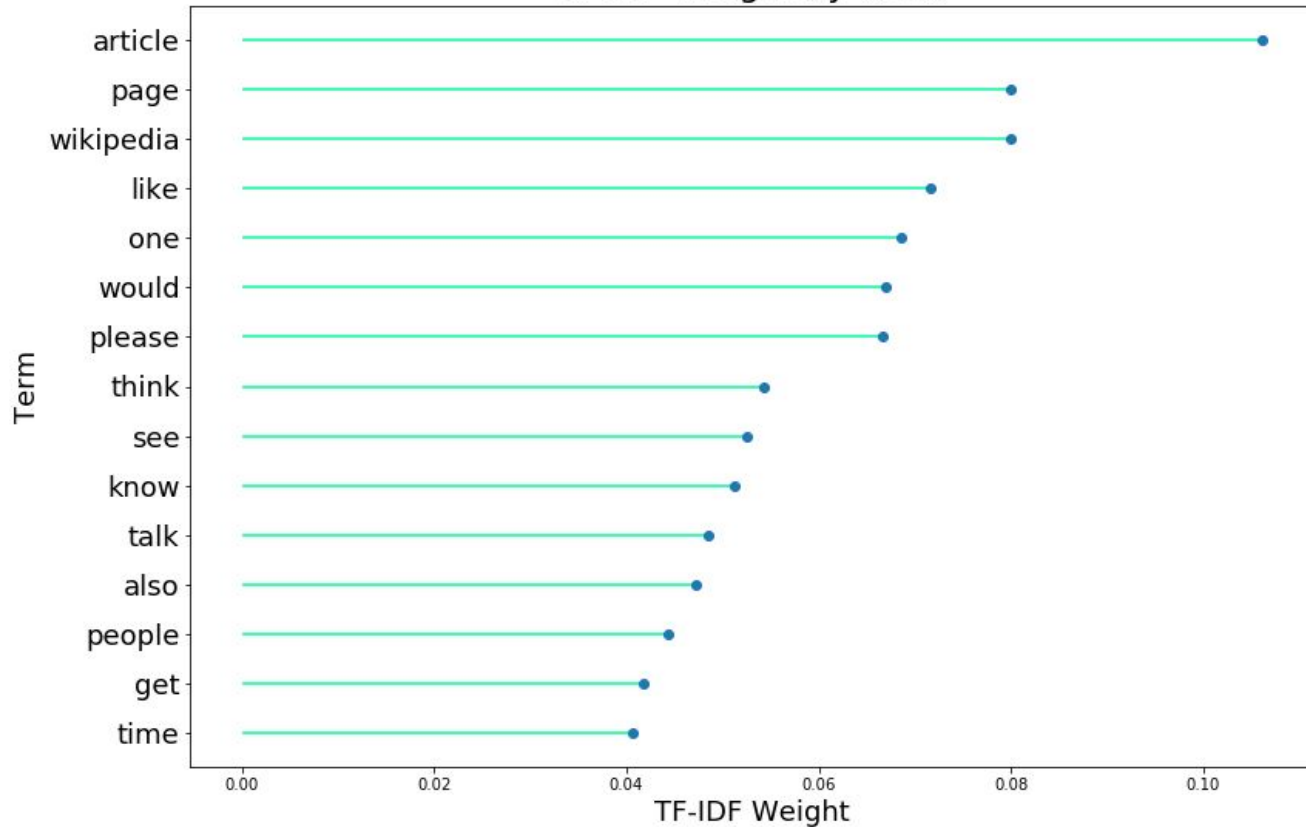
Thank you for your
time today! Please
feel free to connect
with me!



A photograph of a person's hands typing on a laptop keyboard, overlaid with a teal and purple gradient. The person is wearing a light-colored top and a ring on their finger. A white mug is visible on the desk in the background.

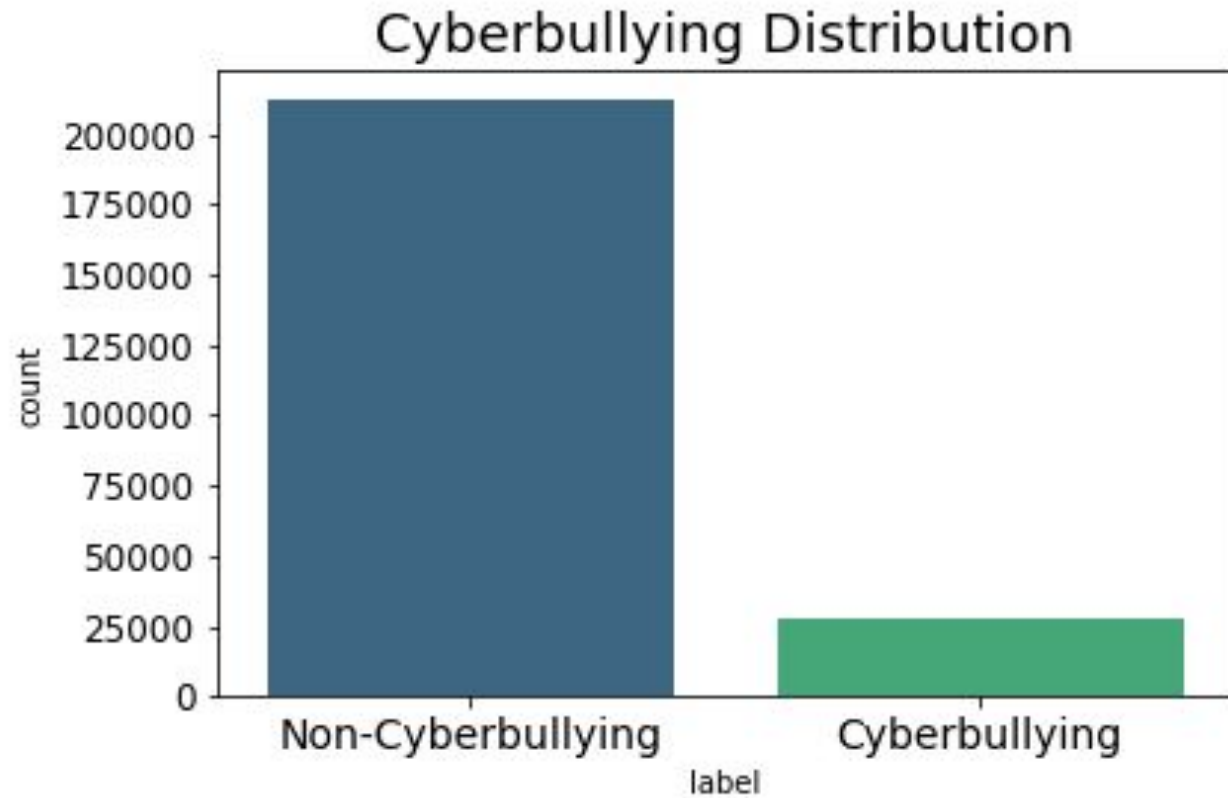
| appendix

TF-IDF Weight by Term

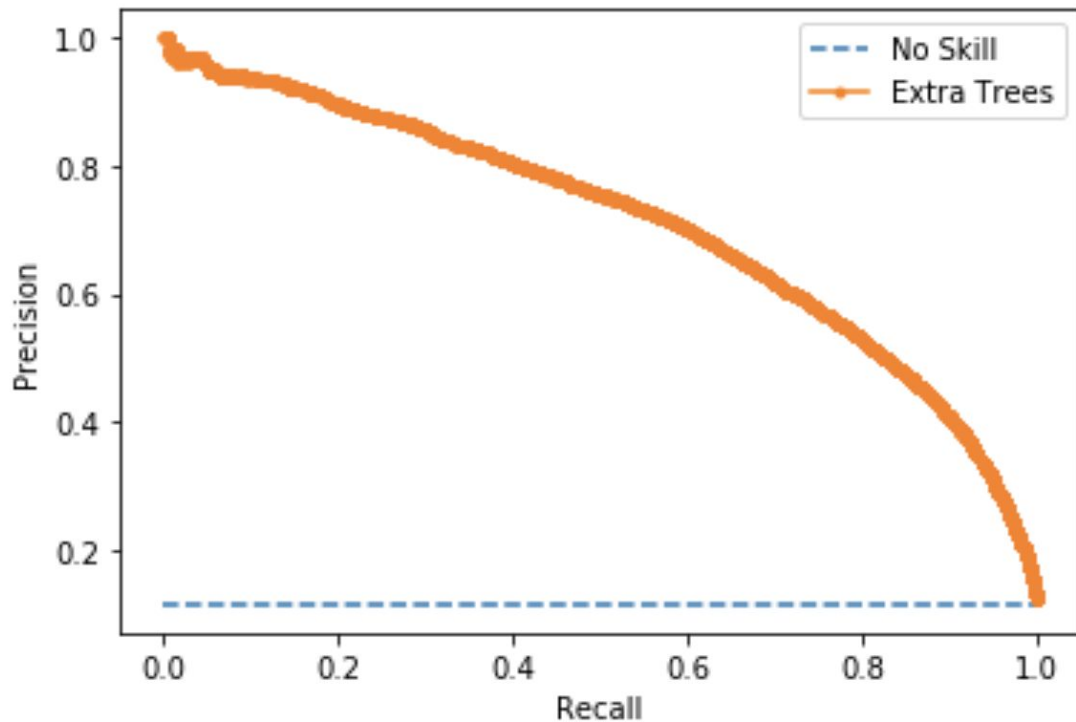


appendix

appendix



F1=0.619, AUC=0.710



appendix