Formosan Cyber Watchdog

Group 4: Chu-Yun Hsiao, Shih-Yun Lin, Yi-Chun Huang, Yu-Hui Lin

12 / 06 / 2023

Team members




Chu-Yun Hsiao	Yi-Chun Huang	Yu-Hui Lin	Shih-Yun Lin

Table of contents

Craigslist

Classified AD

jobs, apartments, garage sales,
used cars, personal ads etc.

Local focus

Reducing long-distance shipping cost
Community connection

User-friendly System

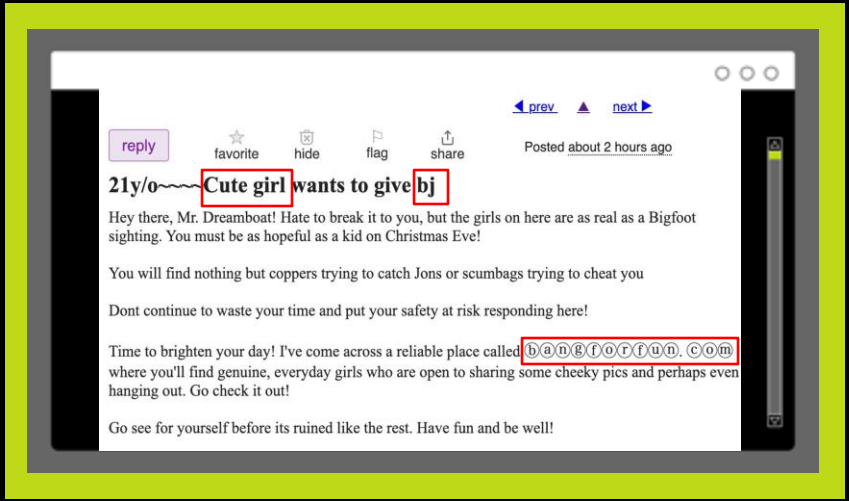
So easy, your child can do it

208 million

visits per month – as of August 2023

Dark side

Robbery
Sexual Assault
Scammer
Child Trafficking and Slavery



Inappropriate content?

filter

By Keywords

Fail to filter the special symbols and abbreviations

flag

post removed when receiving a sufficient number of flags from different users

Time-consuming
Manual reviews

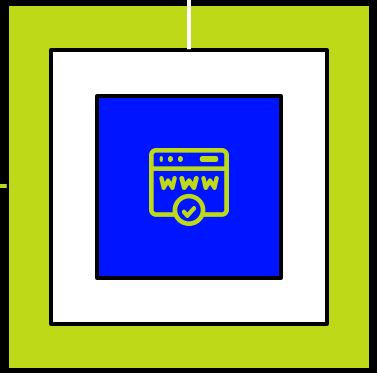
Business Analysis

- Enhancing User Trust and Confidence
- Ensuring a Safe User Experience
- Compliance with Regulations and Policies
- Improving User Engagement
- Boost Advertiser and Partner Relations

Considerations

Scalability & Performance

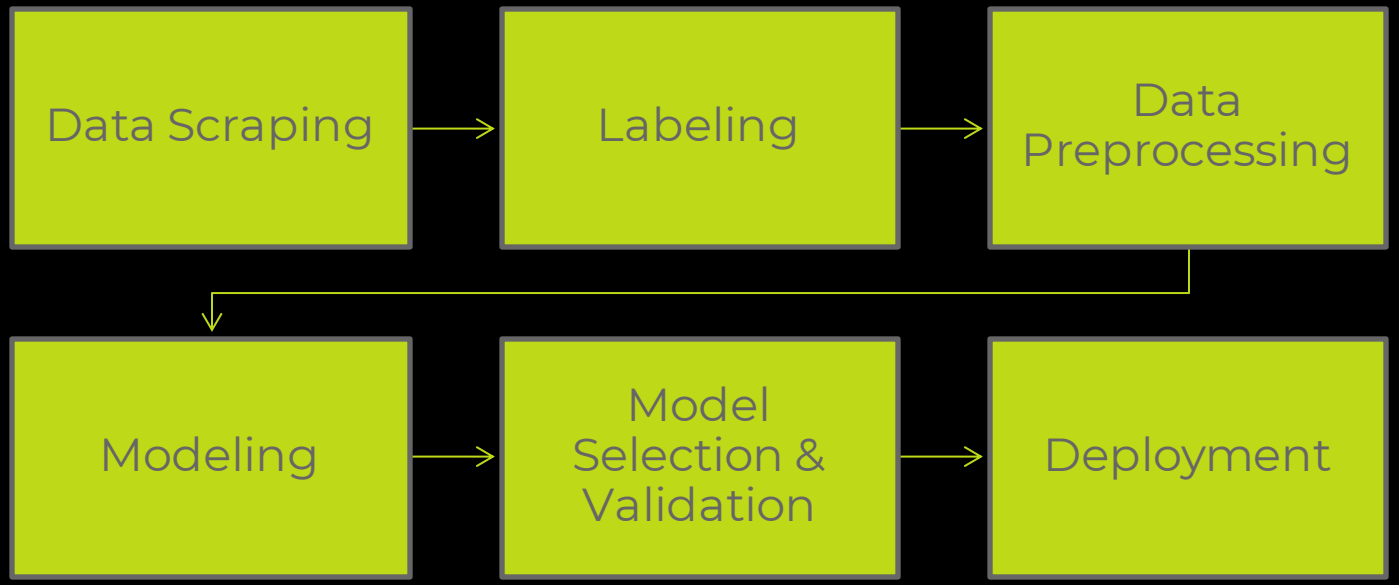
Find the balance between restricting and website volume



Ethical Implications

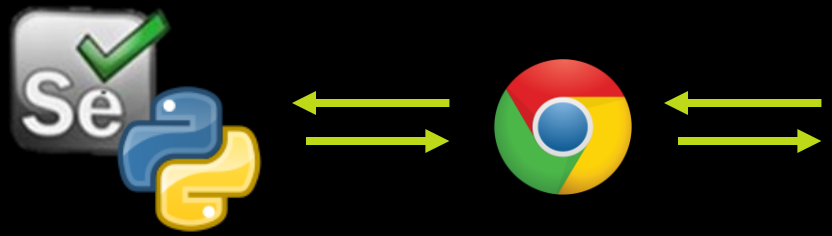
Collect diverse datasets to train the model

Data Analysis Procedure



Data Scraping

- Utilizing Selenium Library to automate a web browser, execute JavaScript, and collect data
- 11,000 Posts on Craigslist's Community Section



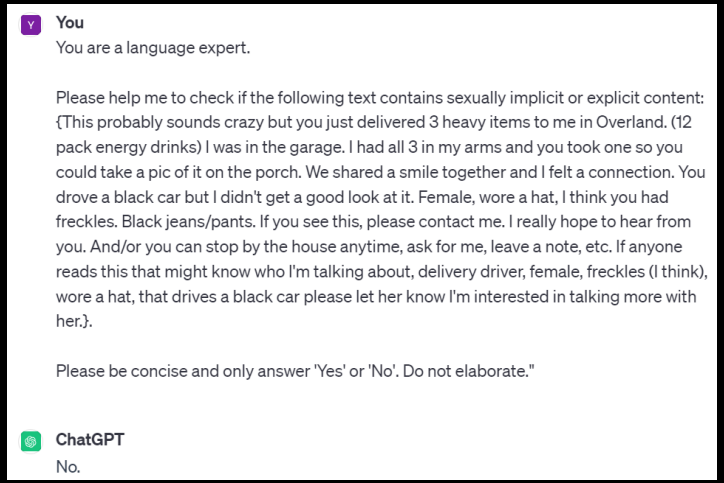
craigslist

community

activities	lost+found
artists	missed connections
childcare	musicians
classes	pets
events	politics
general	rants & raves
groups	rideshare
local news	volunteers

Data Labeling

- Label prediction: supervised learning
- Applied LLM(Large language models) to label posts
- Manually tried different models, Google: PaLM 2 Chat 32k, OpenAI: GPT-3.5, Mistral 7B, Zephyr 7B, etc.
  - Decision: Google PaLm2 Chat 32K
- Sent prompts through API and labeled posts based on the responses



Data Prepossessing

- Tokenizing, lemmatizing, and removal of punctuations, stopwords, and numbers.
- Word-embedding: TF-IDF, 1&2-gram, minimum 20 times
- 3,720 words(in 1- or 2-gram) as independent variables
- Data partition: 70% training set & 30% test set

Modeling

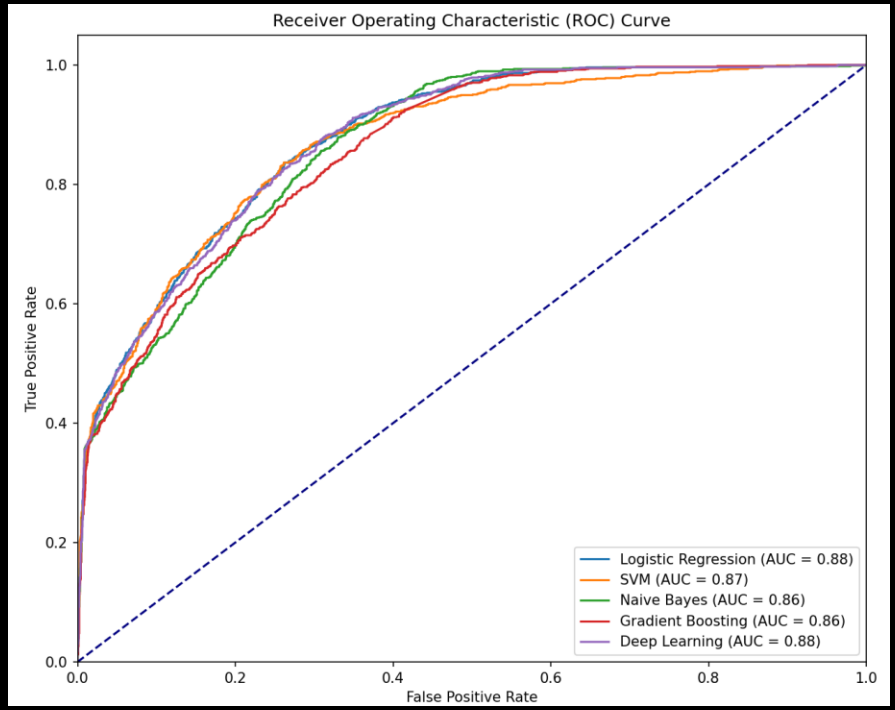
- Y = f(X1, X2, X3...), two-class classification
  - Appropriate vs. Inappropriate
- Logistic regression
- Naïve Bayes and SVM
- Gradient boosting
- Neural network

	Variable	Coefficient
1265	girl	2.676162
1190	fun	2.593670
400	called	1.975589
1069	find	1.955267
1767	let	1.928546
...	...	...
533	christmas	-1.998285
2993	show contact	-2.044082
2587	puppy	-2.328106
209	band	-2.407132
2715	rehoming	-2.509852

	Feature	Importance
400	called	0.129758
2427	place	0.097166
2658	real	0.044260
1190	fun	0.036258
3571	wasting	0.035527
2992	show	0.027591
2900	see	0.026709
1486	home	0.024740
502	chick	0.023192
1359	guy	0.020294
2715	rehoming	0.019537
3642	woman	0.018684
1027	fee	0.018607
1265	girl	0.018606
2993	show contact	0.016047
2455	please	0.013823
21	actually	0.011650
209	band	0.010891
180	available	0.010033
3308	think	0.009944

Model Selection

Model	Criteria: AUC
Logistic Regression	87.8%
Naive Bayes	86.5%
SVM	87.0%
Gradient Boosting	85.9%
Neural Network	87.6%



Prediction Example:

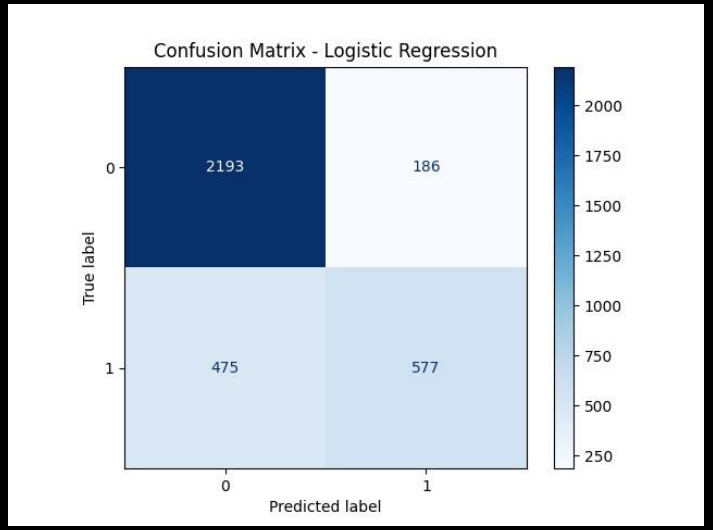
Demo1	Just ⒼOOⒼLⒺ the name 🅱🅰🅽🅶🅵🅾🆁🅵🆄🅽. 🅲🅾🅼 and you can view it right in the top While you are ready to stop wasting time, move see for your own. You will be cheerful you did!
Demo2	Don’t risk your job, relationship, or freedom replying to these posts! There's still one sight that hasn't been tainted by cheaters or the feds. It's called ⓅⓊⓈⓈⓎⒷⒾⓏ and it's a haven for everyday girls who are just looking for some company
Demo3	Older Women (Silver Spring) Mid 40’s SBM loves the company of women 50+. Extra points if unshaven
Demo4	Pen pal (Akron) Older guy looking for someone to chat with. In my mid 60s and would prefer someone around my age. And preferably a female. Easy to talk to, non judgmental. Someone ?

Prediction Demo

```
Logistic Regression Predictions for new samples: ['1' '1' '0' '0']
SVM Predictions for new samples: ['1' '1' '0' '0']
Naive Bayes Predictions for new samples: ['1' '1' '0' '0']
Gradient Boosting Predictions for new samples: ['1' '1' '0' '0']
Deep Learning Predictions for new samples: ['1' '1' '0' '0']
```

Confusion Matrix(Classification):

- Type 1 error (Not harmful but predicted as harmful): 186/(2193+186)≈7.82%
- Type 2 error (harmful but predicted as not harmful): 475/(577+475)≈45.15%



Threshold(0~1):

- Tradeoff between user experience(Type1 error) and restriction level
- By creating a UX, adjust threshold to decrease Type1 error(avoid too much intervention).

	Politics	Activities (children-related)
Threshold (Delete)	0.9	0.8
Threshold (Alert)	0.7	0.6

Application

Pop-up Alert Window

Protect children from inappropriate content

_ □X

Adults only (18+)

This website contains
adult material and is only suitable
for those 18 years or older.
Click Enter only if you are at least
18 years of age.

Enter or Exit

Conclusion

Accuracy

New Models: improve the accuracies of the predictions.