OBSCENITY BLOCKER

Project Aeros

कवच
KAVACH
2023
साइबर सुरक्षा हैकथॉन
Cybersecurity Hackathon

Scope of the Project

Finding a technological solution for identifying and blocking any obscene media.

Unsolicited Imagery to Women

Women are often harassed with profane language or obscene content.

Unsafe Content for Children

Kids are exposed to unsafe content on the Internet at a young age.

Blocking in Professional Settings

Professional environments require certain words or media to be blocked.

Solutions

Ways that our model solves this problem.

Detecting Profane Language

Identifying cuss words in text and detecting as profane.

Detecting Profane Audio

Performing speech-to-text analysis to detect cuss words.

Identifying Images as Obscene

Distinguishing images as obscene/non-obscene using image classification.

Target Market

This model applies to every user on the Internet who may need to use an obscenity/profanity blocker.

However, the vulnerable stakeholders are:

Women

Women are exposed to unsolicited media from harassers online.

Children

More and more children are frequently on the Internet, a highly unsafe platform.

Model Implementation

Source code of how this solution is actually implemented.

Training model using kernel SVM classifier

```
from sklearn.svm import SVC
classifier = SVC(kernel = 'rbf', random_state = 0)
classifier.fit(X_train, y_train)
```

```
SVC(random_state=0)
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.

On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

```
y_pred = classifier.predict(X_test)
accuracy_score(y_test, y_pred)
```

```
0.9453934095494284
```

Produces an accuracy of 94.5%

```
testtext="you are not someone I would like to talk to "
df=cv.transform([testtext]).toarray()
df=clf.predict(df)
if (df==1):
  print("profanity detected")
else:
  print("profanity not detected")
```

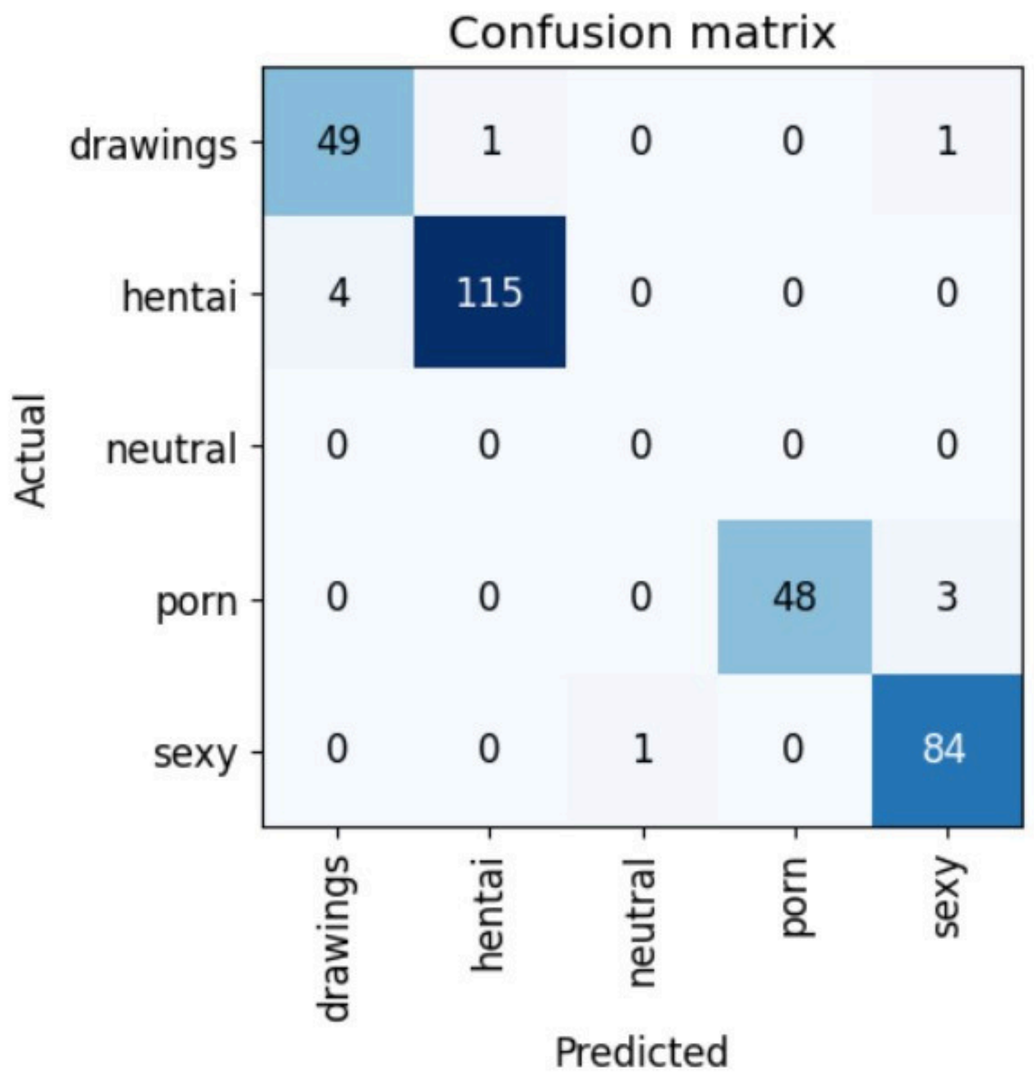
```
profanity not detected
```

```
testtext="I hate you ,fuck off  "
df=cv.transform([testtext]).toarray()
df=clf.predict(df)
if (df==1):
  print("profanity detected")
else:
  print("profanity not detected")
```

```
profanity detected
```

Test Cases

```
interp = ClassificationInterpretation.from_learner(learn)
interp.plot_confusion_matrix()
```



Model Roadmap

- Detection and classification of obscenity
- Developing an API
- Smooth integration to all apps of the user

Model Demonstration

How this model will be deployed and
will look to the user.

Future Scope-

How the model could further be developed.

Block the media from being uploaded.

Prevent display of sites containing such content.

Report the user to required agency.

Parental Supervision

Giving parents access/alerts to the content that their child is viewing.

Model detects obscene websites being viewed on the child's device

An automated message is sent to the parent's contact.

Business or Revenue Model

Our services could be lent to private stakeholders.

Professional websites like LinkedIn or other corporate companies that can have an in-built blocker of obscene content from showing/being uploaded.

Video classification can be extended to classifying movies as well- providing a censorship rating using a Machine Learning model.

Meet our Team

Aryan Kamani

Daksh Agiwal

Aditi Kulkarni

Bhuvan Noone

Tareen Khan

Srujana Akella

Thank You