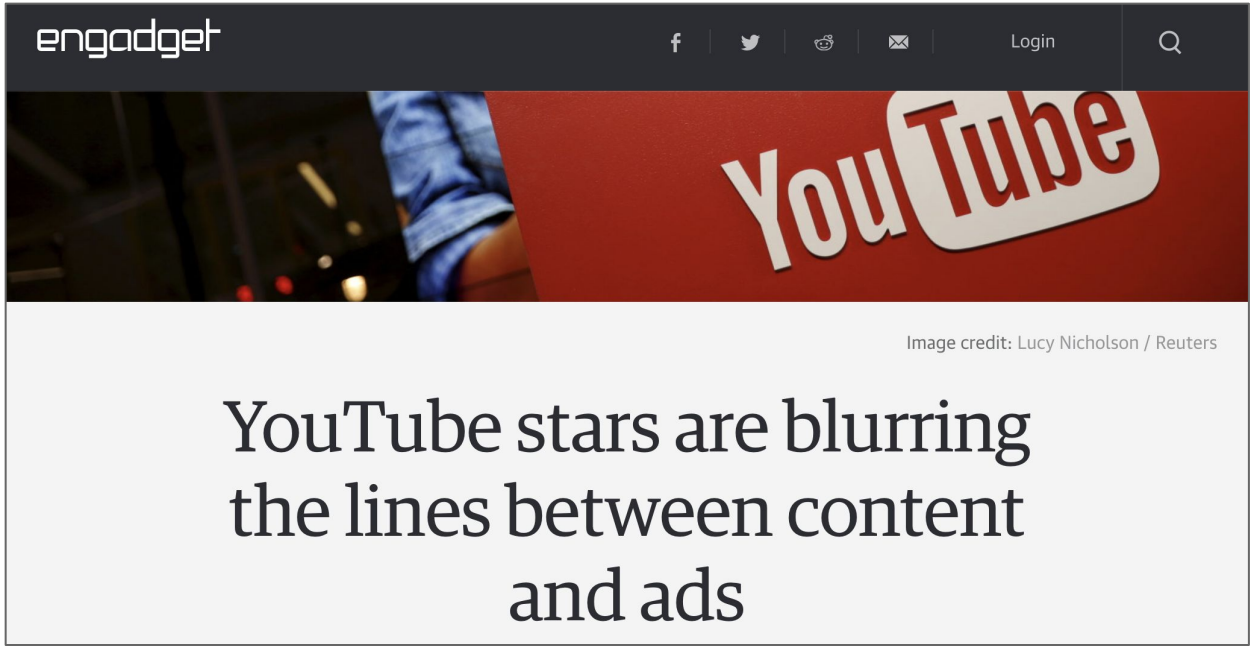
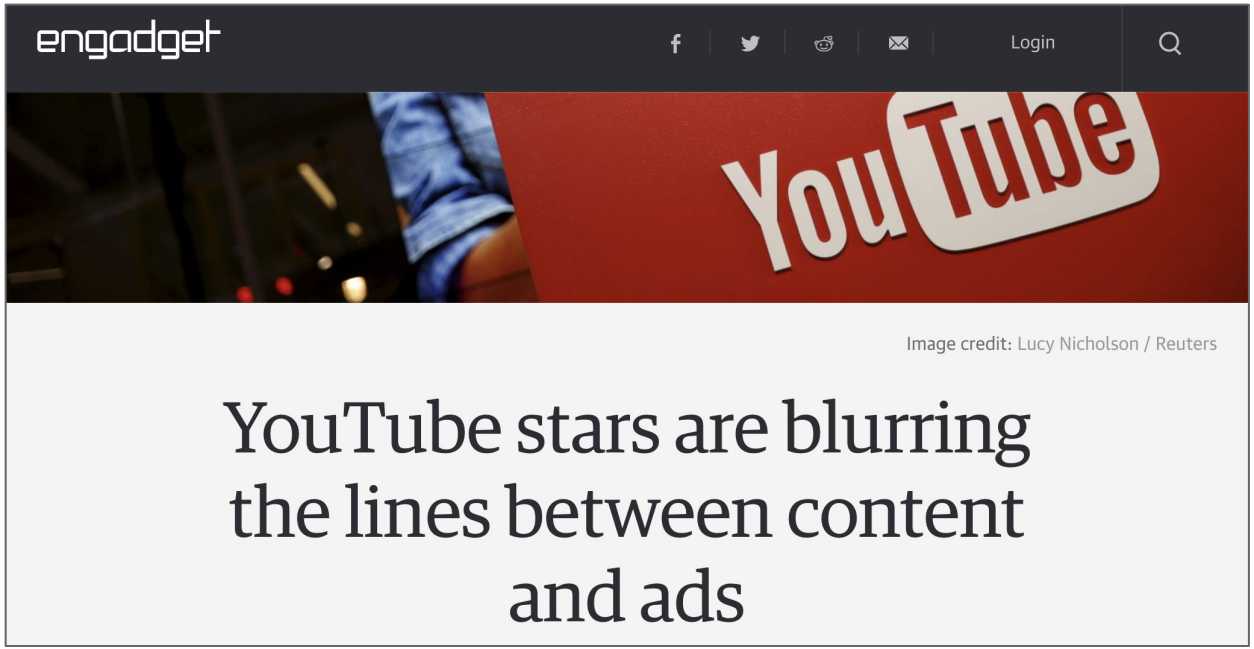
April Harry

Identifying sponsored content in YouTube videos

YouTube content creators partner with brands for **ads**



Some **fail to disclose** sponsored content, against YouTube's Terms of Service

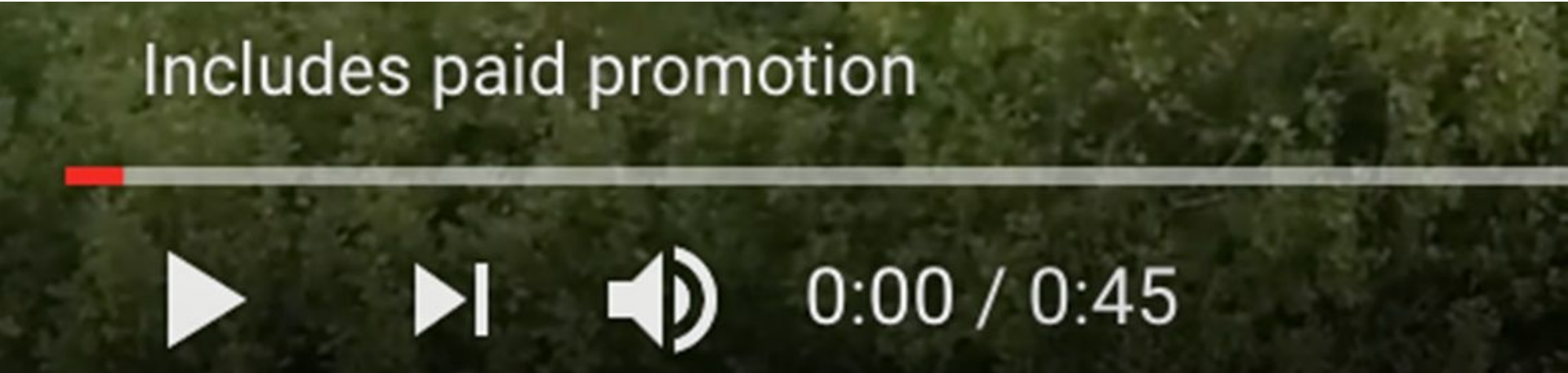


Mislead subscribers out of **fear of losing views**

90 influencers warned by FTC in 2017

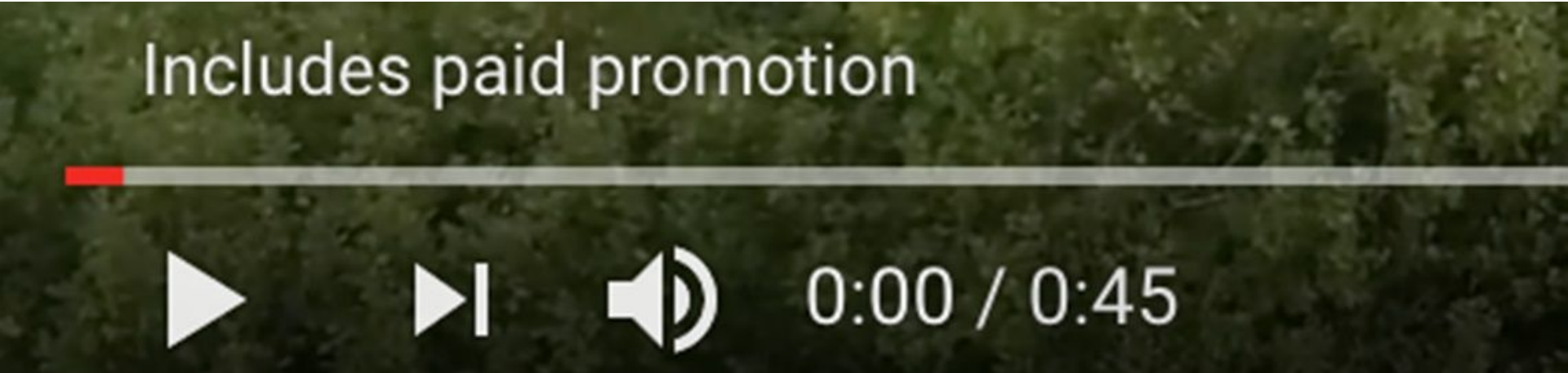
Up to **$40,000 fine**

New feature makes it easier to **disclose sponsorship**



New feature makes it easier to **disclose sponsorship**

...but these ToS **violations** still occur



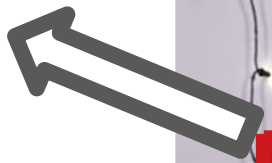
How to **identify sponsored videos**?

www.sponview.site

Scraped auto-generated captions from YouTube videos

BeautifulSoup

youtube-dl



Processed captions using

and

"Hello and welcome to my fishing channel, you are welcomed."

Processed captions using

and

"Hello and welcome to my fishing channel, you are welcomed."

Remove **Stop Words**

"hello welcome fishing channel welcomed."

Processed captions using

and

"Hello and welcome to my fishing channel, you are welcomed."

"hello welcome fishing channel welcomed."

Stemming

"hello welcom fish channel welcom."

Processed captions using

and

"Hello and welcome to my fishing channel, you are welcomed."

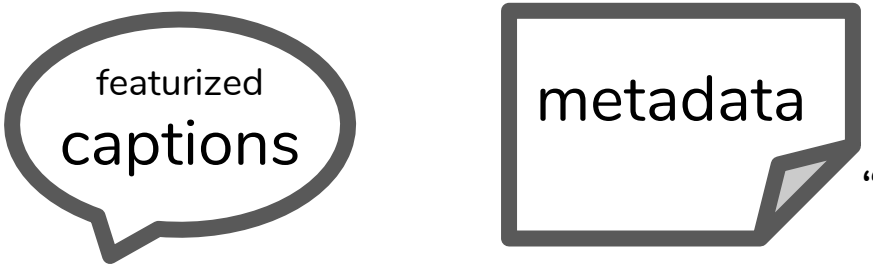
"hello welcome fishing channel welcomed."

"hello welcom fish channel welcom."

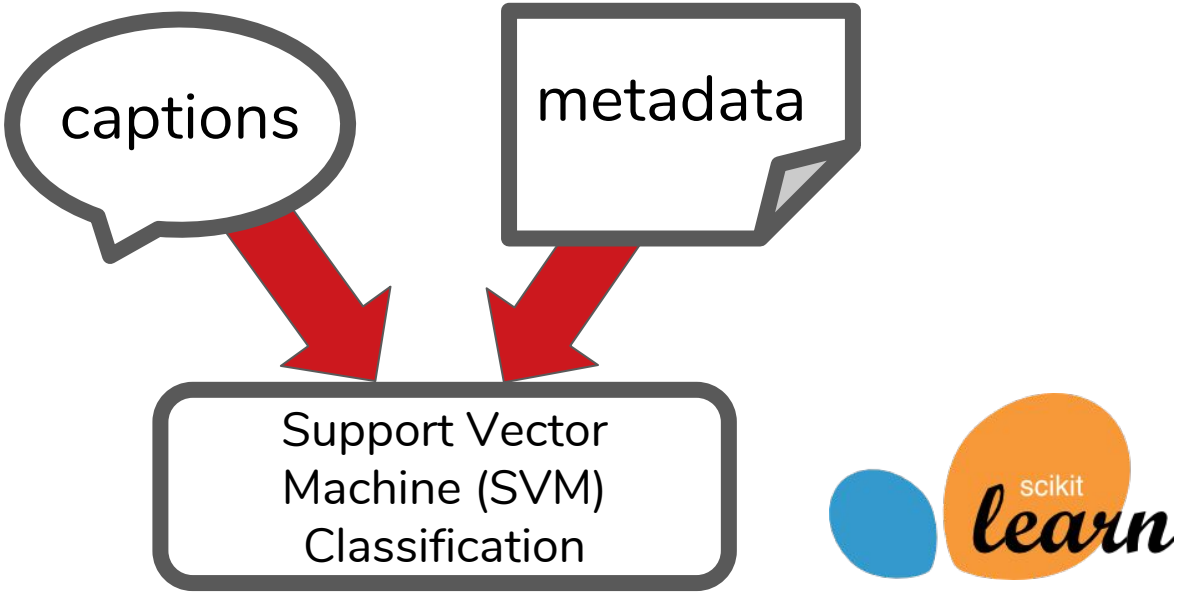
TF-IDF Weighting

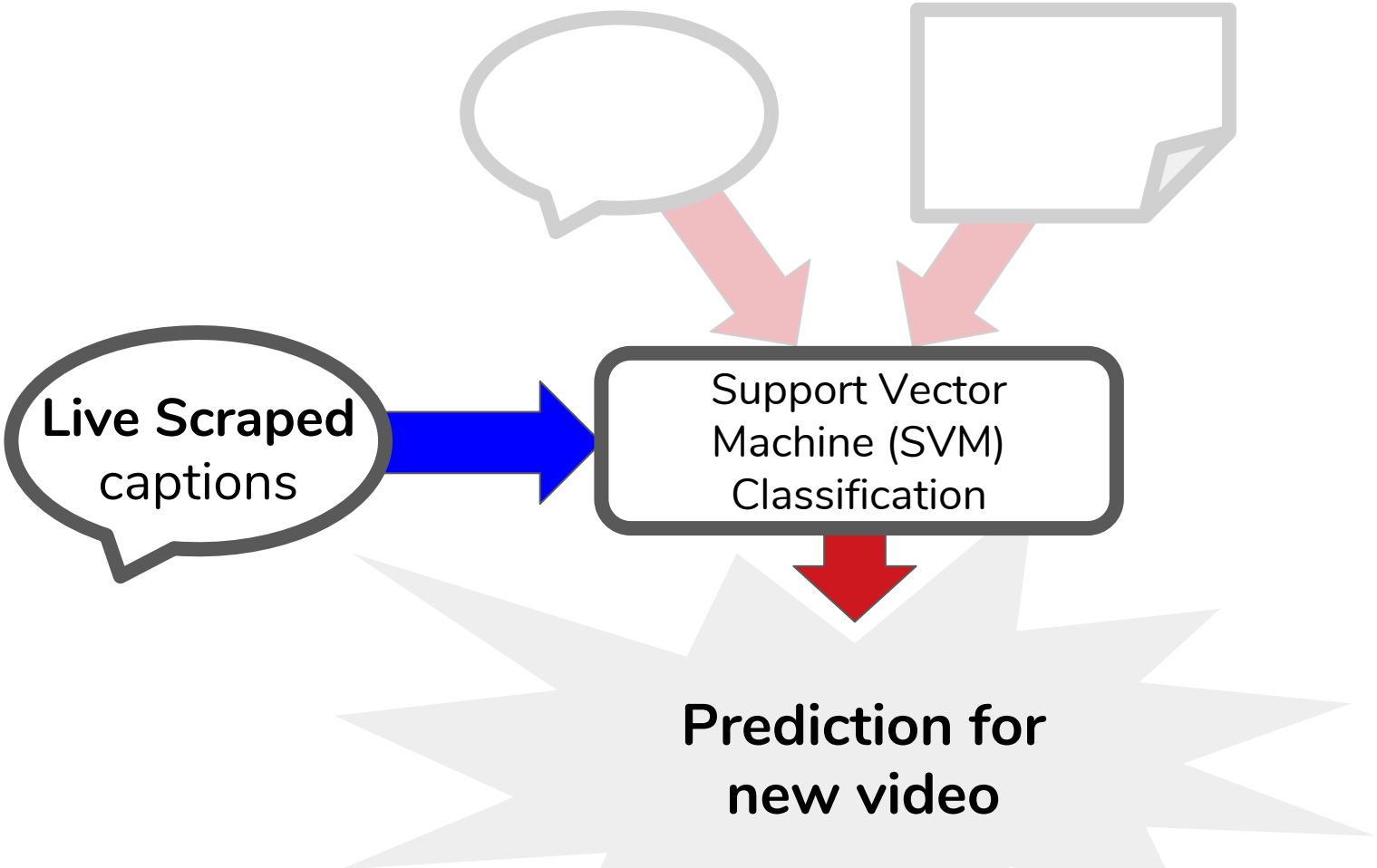
"hello welcom fish channel welcom."

Scraped + YouTube API



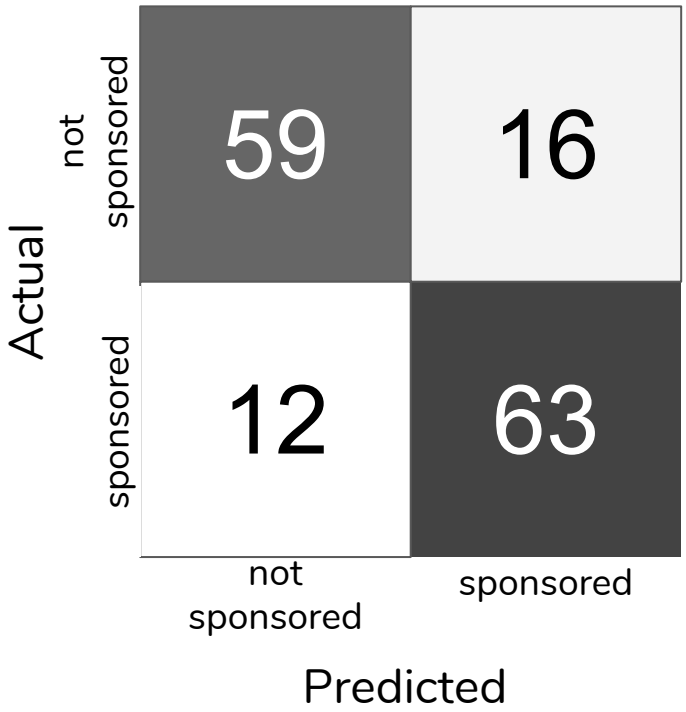
Extract Labels
“**IS/IS NOT** Sponsored”


captions
metadata
Support Vector Machine (SVM) Classification
scikit learn

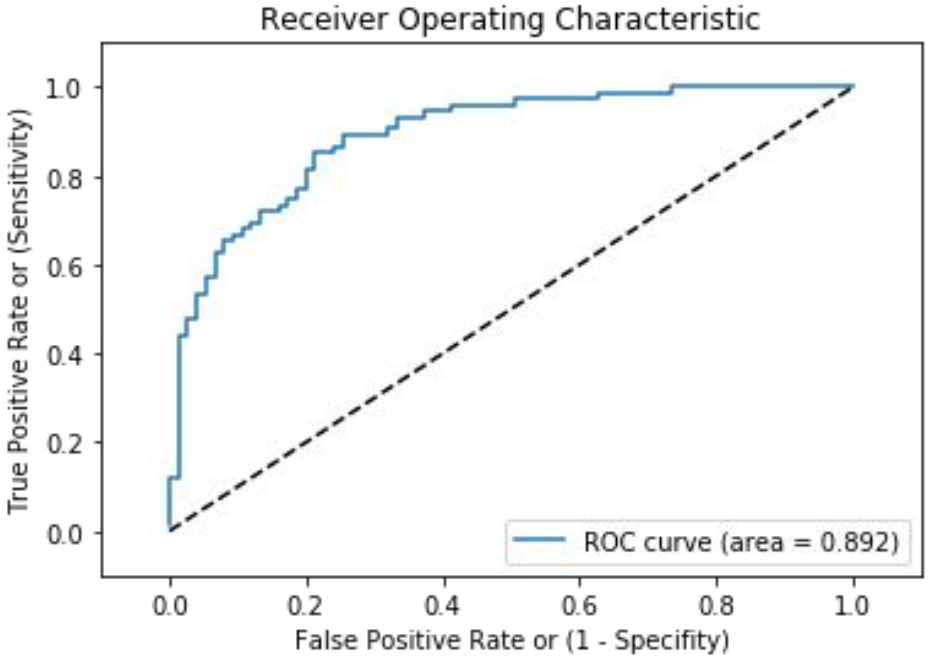

Live Scraped captions
Support Vector Machine (SVM) Classification
Prediction for new video

Predicts with **81%** accuracy on test set

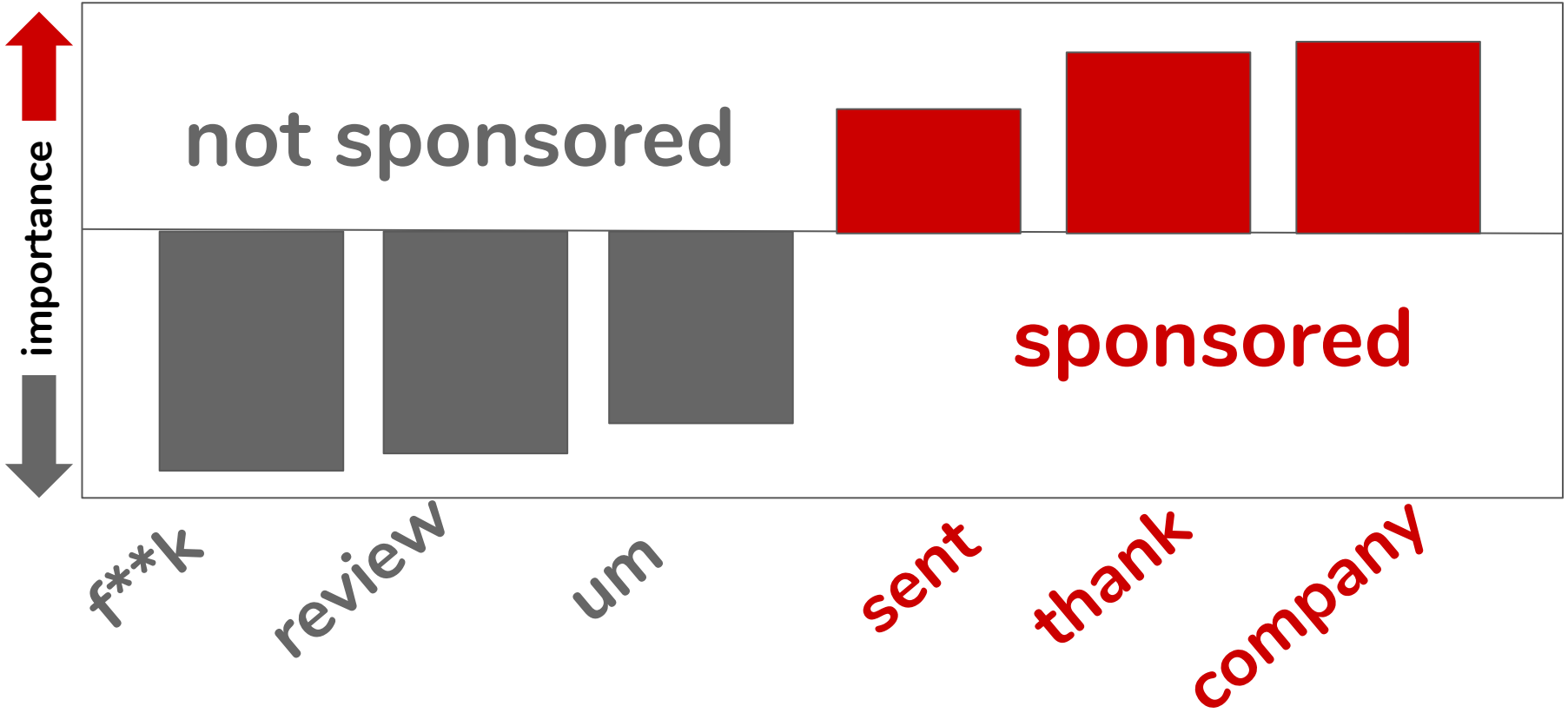
Confusion Matrix



Area Under ROC = 0.89



Words characteristic of sponsored videos



YouTube
identify Terms of Service violations

Users
make informed viewership decisions

FTC
enforce consumer protection laws

April Harry

PhD in **Statistics** from

PURDUE
UNIVERSITY®