



CAN MACHINE LEARNING DISTINGUISH ETHICAL, UNETHICAL, AND ILLEGAL CONTENT ONLINE, AND WHAT DOES THAT MEAN TO EDUCATORS

GG BERG, RESEARCHER, UNIVERSITY OF SAN FRANCISCO

www.usfca.edu



STUDENTS' CHANGING ATTITUDES

Science News

from research organizations

Student attitudes toward cheating may spill over into their careers

Date: November 27, 2019

Source: San Francisco State University

Summary: A new study finds that students who tolerate cheating in the classroom may also turn a blind eye to unethical behavior in the workplace.

"If [students] have this attitude -- that it's OK to cheat in school -- that attitude unfortunately will carry over to the corporate boardroom," Foo Nin Ho, SF State Professor and Chair of Marketing

In addition ...

The results revealed that group-oriented students, or collectivists, had a more laissez-faire attitude toward cheating than individualistic classmates.

**BIG implications for student's
online social profiles**

BUSINESS' INCREASING SOPHISTICATION

- Over 70% of companies use social media checks to screen job candidates(2018)

SOCIAL MEDIA CONTENT THAT CAUSES EMPLOYERS TO ELIMINATE CANDIDATES

POSTED PROVOCATIVE OR INAPPROPRIATE PHOTOGRAPHS, VIDEOS OR INFORMATION	40 percent
POSTED INFORMATION ABOUT THEM DRINKING OR USING DRUGS	36 percent
HAD DISCRIMINATORY COMMENTS RELATED TO RAGE, GENDER, RELIGION, ETC.	31 percent
WAS LINKED TO CRIMINAL BEHAVIOR	30 percent

Can Natural Language Processing and modeling distinguish the nuances between unethical and illegal behaviors online

BIG implications for students' future jobs and careers

MODEL ETHICAL, UNETHICAL AND ILLEGAL BEHAVIORS POSTED ONLINE



front page of the internet

Reddit is a network of communities based on people's interests. Members find communities they're interested in, and become part of an online community!

Members/ posters are self selecting

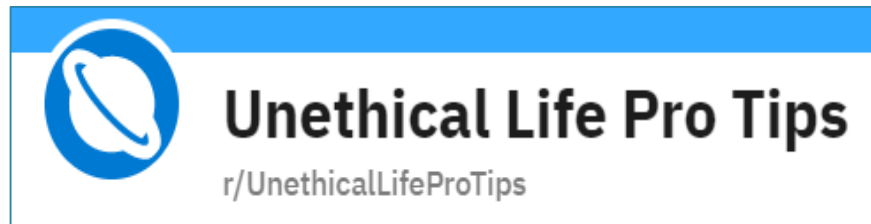
Subreddits are a good source of categorical data

DATA SOURCES



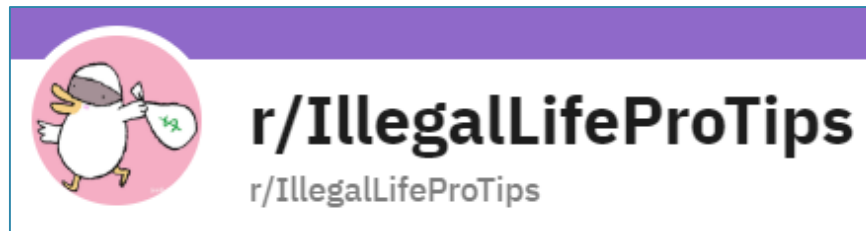
Tips that improve your
life one way or another.

7,613



... perhaps at the
expense of others ...

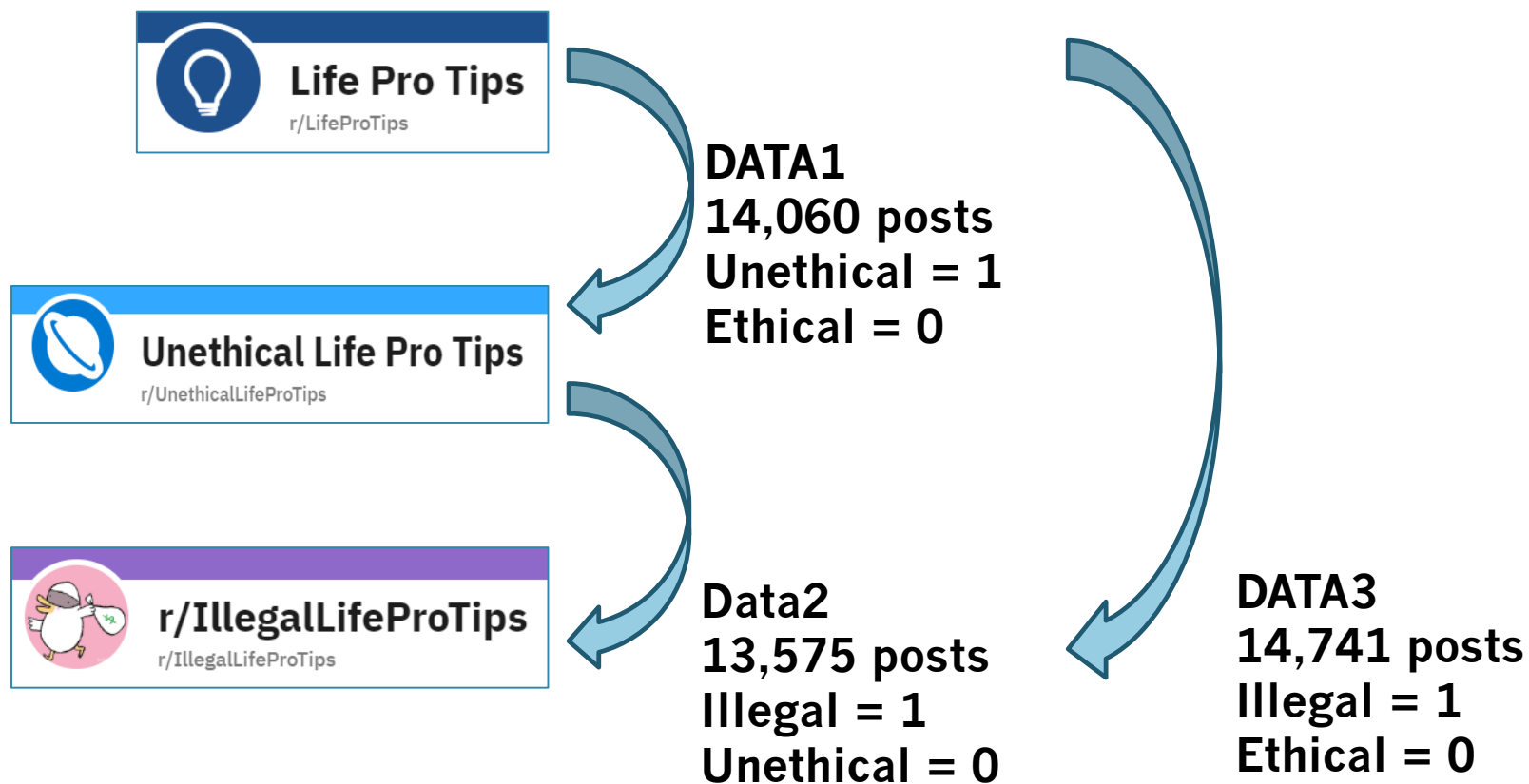
6,447



... illegally

7,128

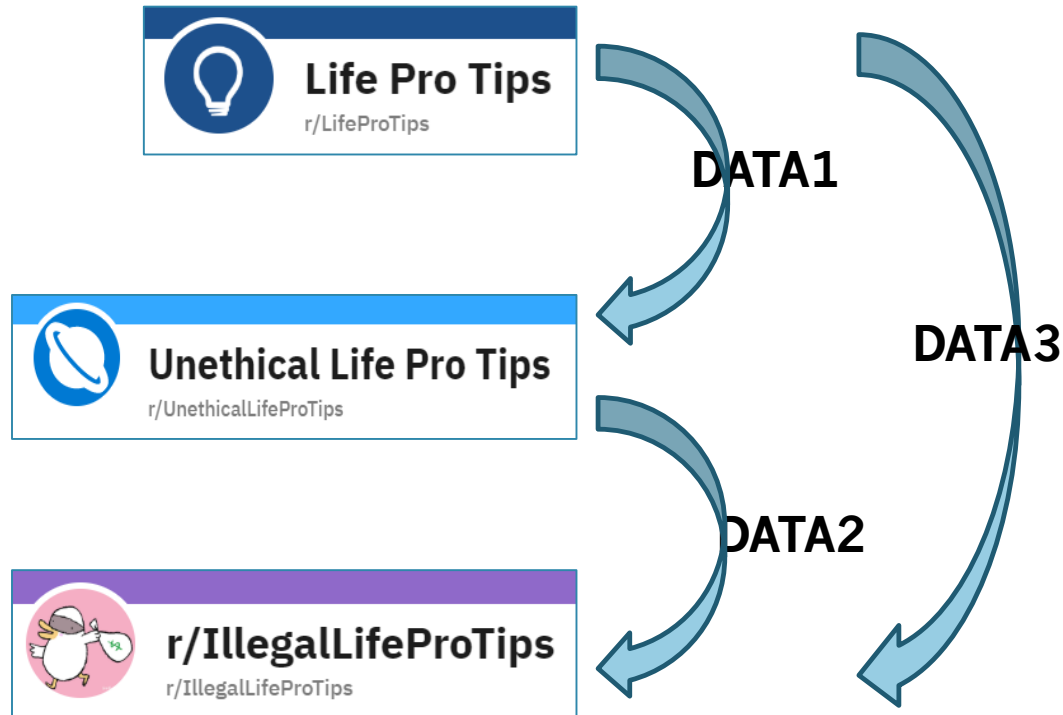
MODELING 'ETHICAL', 'UNETHICAL', AND 'ILLEGAL' LIFE PRO TIP SUBREDDITS



Modelling techniques:

- CountVectorizer
- TfidfVectorizer
- Logistic Regression
- NB Binomial/ Multinomial
- Support Vector
- Decision Tree

INITIAL RESULTS



- **Overall results for all 3 data**

- Titles more accurate than selftext
- Tvec more accurate than Cvec
- Stopwords not an influence
- LogReg most accurate model!
- Multinomial Naïve Bayes slightly less accuracy than LogReg but consistently higher Specificity
- Support Vector similar accuracy to LogReg, but with higher overfit and HUGE Time/CPU cost.
- Decision Tree slightly less accuracy than LogReg and highest overfit.

Stopword ethical is found 7 times in Data3 title
Stopword unethical is found 3 times in Data3 title
Stopword legal is found 153 times in Data3 title
Stopword illegal is found 86 times in Data3 title
Stopword evil is found 5 times in Data3 title
Stopword bad is found 107 times in Data3 title

BEST MODELING RESULTS

Dataset	Base	Best model	Results*
Data1 Unethical Vs Ethical	54.1%	TfidfVectroizer 'tvec__max_features': 1500, 'tvec__min_df': 1, 'tvec__ngram_range': (1, 1), 'tvec__stop_words': None LogisticRegression	Accuracy =78.2% Sensitivity =71.5% Specificity =83.8%
Data2 Illegal vs Unethical	52.5%	TfidfVectorizer 'tvec__max_features': 1200, 'tvec__min_df': 2, 'tvec__ngram_range': (1, 1), 'tvec__stop_words': None LogisticRegression	Accuracy =76.0% Sensitivity =80.0% Specificity =71.7%
Data3 Illegal vs Ethical	51.6%	TfidfVectorizer 'tvec__max_features': 2000, 'tvec__min_df': 2, 'tvec__ngram_range': (1, 1), 'tvec__stop_words': None LogisticRegression	Accuracy =89.8% Sensitivity =89.2% Specificity =90.4%

* MultinomialNB very similar results, but with much higher SPECIFICITY

CONCLUSIONS



■ Conclusions

- Simple modeling yields reasonable accuracy in predicting if a post/ behavior is unethical or Illegal
- More data and more modelling will increase accuracy.
- Reasonable to assume employers are capable of screening students / job candidates based on nuanced language.

■ Next steps

- Expand on current study to : 1) ensure modelling accuracy, and 2) refine how online groups influence students' 'collectivism'
- Advise University Career Services to enact stricter guidance to graduating students about cleaning social profiles.
- Approach University staff to urge investigation on including the subject of Ethics in required curriculum



THANK YOU