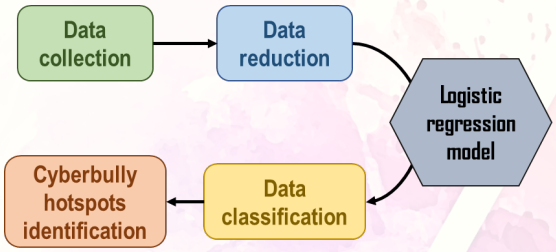
# Detecting Cyberbullying "Hotspots" on Twitter: A Predictive Analytics Approach

Shuyuan Mary Ho, Wenyi Li (Florida State University, USA)
Dayu Kao, Ming-Jung Chiu-Huang, Chung-Jui Lai
(Central Police University, Taiwan)

## Abstract

Predictive analytics modelling provides preventative confirmation to mitigate psychological harm and protect against victimization by identifying cyberbullying "hotspots" on social media.

## Study Framework




## Data Analysis

### *Logistic regression model*

*DV = Cyberbullying (1=positive, 0=negative)*
*IV (predictor variables) = 10 features below*

| Variables | Coef. Estimate | Std. Error | Z-value |
|---|---|---|---|
| Intercept | -4.270 | .321 | -13.308*** |
| You | .007 | .025 | .283 |
| Negative emotion | .127 | .030 | 4.287*** |
| Anger | .006 | .041 | .150 |
| Biology | -.235 | .065 | -3.633*** |
| Body | .281 | .054 | 5.166*** |
| Health | .296 | .071 | 4.179*** |
| Sexual | .428 | .068 | 6.308*** |
| Ingestion | .248 | .077 | 3.215*** |
| Death | .496 | .060 | 8.328*** |
| Swear | .169 | .034 | 4.943*** |

### *Predictive model:*

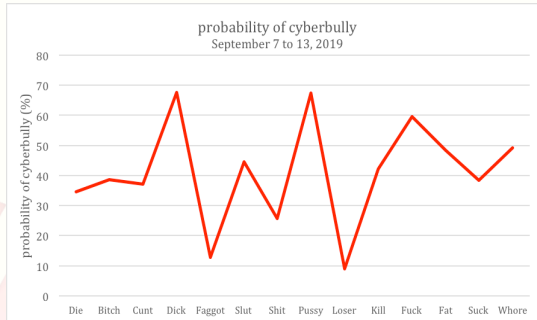
$$\log\left(\frac{P(y=1)}{P(y=0)}\right) = -4.270 + .007you + .127negemo + .006anger - .235bio + .281body + .296health + .428sexual + .248ingest + .496death + .169swear$$

*Classification results in comparison with Kasture (2015)*

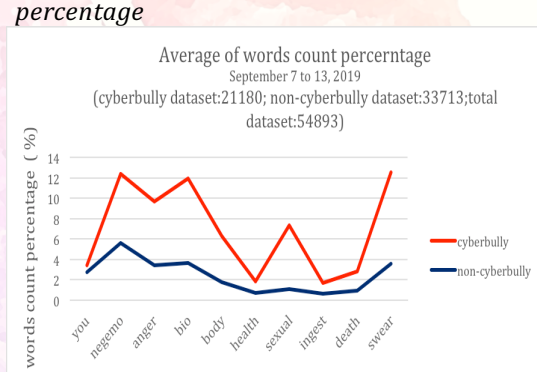
| Datasets | Time-lines | Cyberbullying | | Non-Cyberbullying | |
|---|---|---|---|---|---|
| | | Tweets | percentage | Tweets | percentage |
| Kasture (2015) data | 2015 | 376 | 28.64% | 937 | 71.36% |
| Our dataset | 09/07-09/13/2019 | 21,042 | 38.33% | 33,852 | 61.67% |

## Results and Findings

*Probability of cyberbully using charged language-action cues*




*Comparison of the average of wordcount percentage*




*Predictor Variables w/ large Cohen's d effect size*

| Category | Cohen's D Effect Size |
|---|---|
| You | .083(negligible) |
| Negative emotion | .830(large) |
| Anger | .865(large) |
| Biology | 1.100(large) |
| Body | .816(large) |
| Health | .300(small) |
| Sexual | 1.067(large) |
| Ingestion | .309(small) |
| Death | .438(small) |
| Swear | 1.248(large) |

## Conclusion & Contribution

The study contributes a predictive analytics approach to detecting cyberbullying "hotspots," by identifying the tendency for cyberbullying based on charged language-action cues.