Multi-Angled Statistical Approach to Human Trafficking Detection & Profiling
Yeng Saanchi, Marshall Wang, Saran Ahluwalia, Eric Laber, Sherrie Caltagirone
Laber-Labs
ysaanchi, lwang31, ssahluwa, eblaber@ncsu.edu

Summary
• Human trafficking is a form of modern-day slavery that affects millions of people. Escort websites are a primary vehicle for selling the services of trafficking victims.
• We use data scraped from two major escort websites to build a statistical model that discriminates between trafficking victims and independent providers. The proposed model exhibits better predictive accuracy than competing methods.
• We created a suite of interactive data visualization and exploration tools to identify potential victims and to inform intervention strategies.

Motivation & Challenges
• Human trafficking ranks as the third most profitable crime and generates 150 billion dollars/year.
• 63% of sex trafficking victims report being advertised online, most commonly on Backpage.1
• Raw data are unstructured text and characterized by nonstandard spelling and grammar and a large number of non-English characters.
• Adversarial environment: Traffickers use slangs, intentional typos, obstructed images, multiple accounts...

Data
• 10K annotated Backpage posts provided by Marinus Analytics LLC.2 Each post is assigned an ordinal score from 0-6 coding the likelihood that the provider in the post is a trafficking victim.
• 170K and counting raw posts scraped from Backpage, Craigslist, and TNABoard.
• Key variables extracted from raw data: Account ID, Location, Phone Number, Price, Age, Age Flags, Trafficking Flags...

Predictive Model
• Gated Feedback Recurrent Neural Network3 that predicts the likelihood that a post is for a trafficking victim from the raw text.
• Achieved state-of-the-art performance compared to results from previous papers:
  - 81.1% binary classification accuracy
  - 76.3% weighted binary classification accuracy
  - 0.799 mean absolute error
• Example of how the model updates the predicted probability word by word in the binary classification case:

Search Tool
• Uses a fast and flexible fuzzy matching algorithm.
• Returns the result based on matching score.
• Example of searching for “young student new in town”:

Movement Map
• Visualize the movement of any subject identified by the predictive model or search tool.
• Example: Movement of different users with similar posts:

Future Work
• Expand annotated dataset & improve predictive models
• Face detection and age prediction
• Image segmentation for body ratio estimation

References
1. A report on the use of technology to recruit, groom and sell domestic minor sex trafficking victims, 2015, THORN.
2. Combating human trafficking with deep multimodal models, 2017, E. Tong et al.