Project

Human trafficking analysis by Informs.org (https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Human-trafficking-analysis) will be the case study for my project. Before choosing which model, and how to collect data; we must first define Human Trafficking and why it's important to combat it. Then, we'll see the challenges that the analytics practitioners face.

Wikipedia defines Human Trafficking "the trade of human for labor, sexual slavery or commercial sexual exploitation for the trafficker or others". Human Trafficking is a crime because "of the violation of the victim's rights of movement through coercion and their commercial exploitation". The typical victims are women and children. Human trafficking is estimated $150 billion in profits in 2014, according to Internation Labour Organization. In 2012, it was estimate that 21 million of the victims, %68 percent was exploited for labor, %22 percent for sex, and %10 for in-state imposed force labor.

.
The challenge deals mostly with the data itself. Because traffickers are covert, and victims are concealed. This poses a real problem for anyone trying to understand and evaluate the data. Some data are missing(or false), the overall dataset can be small, and traffickers will relocate. Hopefully, these challenges can be mitigated with these model and collecting data methods .

To combat missing data, we can try to do data imputation or data substitution. We have various ways impute the data. We can try mean or regression imputation for continuous missing data, or mode to replace missing categorical data. With the caveat that certain data is missing and it should be left missing, because we may have reason to believe that it was intended to be.

One of the main issue facing the law enforcement is the volatility in the data, because victims and traffickers will migrate to elude law enforcement. One way to combat this is to quickly realize that a trafficker has moved to a new location, so time plays a key role. This can be achieved via change detection. Which means that the data must be or transformed to fit sequential/time series . We can transform non-sequential data type, by adding a timestamps to each sample point(in the order that which it was given). Afterwards, we can try autoregressive or moving average model to detect any changes in the time series. We can try CUSUM or volatility clustering with GARCH. GARCH and ARCH models are especially useful models when the goal of the study is to analyze and forecast volatility. So once a data is steps out of the threshold for example, this could alarm law enforcement that victims and traffickers have relocated. The next thing to consider is which factor(s) needs to be further analyzing to have a threshold for and what is the appropriate threshold.

In theory, if we have a solid time series dataset, we could extract interesting information; for example, trends, cycles and seasonality. However, as we state before, we are limited information on both victims and traffickers. Hence, my next proposal is find new ways to gather information.

Another approach to the problem of catching moving traffickers is to plot them on a grid map. We can use various websites/social media to hint us in the right direction. For example, we can use Google analytics(for website traffic) or Google Trend to search for keywords that will give us locations where that keyword was search frequently. Sometimes to find human traffickers/victims, we must locate where the demand is. We'll look for buyers, because it's the buyer's demand that needs to be met. And social media is one way that traffickers communicate with their buyers. These can be dating sites,dating apps, Instagram, Twitter, Facebook,Craiglist etc. So with the new data , we can pinpoint buyers and label them as buyers . Then we can also pinpoint traffickers on the map, and label them as such. Hence after we have enough data points, we can use a proximity search like Nearest neighbor search for example, to get an estimate where possible trafficker and even forecast where may go next.

Last approach may be unconventional, and it's possible that it may not be as helpful. However, sometimes you have to try new things to gather new and interesting facts . Given that we have images of trafficker and trafficked. These images for example can be mug shots of the individuals, it's important

that these image stay uniform. With the images we can do run it through a Convolutional neural network or CNN. Often CNN is recognize for it's uses of image classification, video analysis and natural language process. We'll use it for image classification, and also to extract interesting features that we may have missed. Why would physical appearance matter? Because from what I understand, often A "lover boy" or "romeo" pimp traps his victim and recruits them into the life by maintaining a facade of romantic attraction. From the images maybe we can see if these "pimp" types or "romeo" type have any physical appearance that can be capture/extracted using CNN . If our image data is small, we can try L1 and L2 regularization and pretrained convent to get the most of the small set of images. Also, we can superimpose the class activation heat map on the original pictures, which is a visualization technique to see where it started to classify images. From that point, we as analyst can learn a lot from the activation heatmaps of a convent.

Let's conclude on what are the models and data we thus far to deal with the challenges.

1.Given the current data, fix the missing data point with imputation or leave it missing if necessary.Make sure the data is sequential, Then apply a change detection CUSUM or GARCH and pinpoint when a change in mean or change in variability is occurring. This may reveal when Trafficker/victims have changed location.

2. Gather new data points using various social media apps/websites . With these new data points we can extrapolate them onto a grid map then use clustering methods like Nearest Neighbor to estimate where they're huddling together. This will give us new insight of where they could possibly migrate to.

3. Lastly, we'll you Convolutional neural network for image classification and extract important feature . The idea here is the analyst learn from what CNN see, and make conclusions base on that.