

Human Trafficking Through Data Sharing and Analytics

1st Lam Ying XianAsia Pacific University of Technology & Innovation
Kuala Lumpur, Malaysia2nd Rajasvaran LogeswaranAsia Pacific University of Technology & Innovation
Kuala Lumpur, Malaysia
logeswaran@apu.edu.my

Abstract—Human trafficking remains a defining issue in the 21st century. Analytics in human trafficking is explored to examine the existing methods used to combat the problem. A systematic approach of reviewing the related works is adopted, investigating each element within the definition of human trafficking, analytics techniques adopted by previous studies and the analysis of the related works. The analytics techniques were categorized into machine learning and statistical analysis. Significant findings in this paper include: (1) Most of the previous studies explored sexual exploitation using machine learning and statistical analysis; (2) In machine learning, classification algorithms were widely used; (3) In statistical analysis, data sampling methods were commonly applied; and (4) No machine learning method was found to detect forced labor, which remains an area that needs to be further explored.

Keywords—Human trafficking, sex trafficking, forced labor, analytics, machine learning, statistical analysis

I. INTRODUCTION

In the latest report, the International Organisation for Migration (IOM) estimated that forty million people were trapped in human trafficking in 2016 [1]. It is further predicted that the number of trafficked victims will increase significantly due to the weak socioeconomic situation amid the COVID-19 pandemic. Human trafficking is labelled as the third biggest form of transnational organized crime. Despite joint efforts by the international community, the problem is still pervasive [2]. One of the biggest issues is the difficulty in identifying trafficked victims, which led to the limited data.

Analytics is widely defined in academia. However, the Institute for Operations Research and the Management Sciences (INFORMS) suggested a generally used definition that interprets analytics as “the scientific process of transforming data into insights for the purpose of making better decision” [3]. Having that definition in mind, analytics has great potential to discover the trends in complicated social data and predict future human trafficking activities. Despite the constrained availability of data, numerous literature studies have demonstrated creativity, and leveraged on the online data to track and detect any suspected human trafficking activities.

This paper reviews the efforts undertaken in applying analytics to combat human trafficking. The paper first provides the definition and scope of human trafficking in Section II, followed by an explanation of the adopted methodology in Section III. Section IV presents the analytics methods used, whilst the analysis of the literature studies is given in Section V, with conclusions drawn in Section VI.

II. DEFINITION OF HUMAN TRAFFICKING

The loosely defined term ‘human trafficking’ is confusing, often leading to misidentification. For instance, trafficked victims are frequently misidentified as smugglers, when in fact they are being exploited (US Department of State, 2019).

According to the Palermo Protocol, the internationally recognized definition of human trafficking is categorized into three elements as shown in Fig 1. the act, the means and the purpose [4-6].



Fig. 1. The three elements within the definition of human trafficking

1. The Act of Human Trafficking

A human trafficker is identified if he/she “transfers, recruits, transports, harbors or receives another person for the purpose of exploitation”. The act can take place both domestically and cross borders.

2. The Means of Human Trafficking

Various methods used by traffickers to perform the act of human trafficking were identified, where violence or threats of violence against the victims or their friends and family by the means of ‘coercion, abduction, fraud, deception, abuse of power or giving payment’ fall under the definition of human trafficking.

3. The Purpose of Human Trafficking

The purpose of the exploitation is taken into account in the definition of human trafficking. It includes “sexual exploitation, forced labor, domestic servitude, forced or coerced organ removal and other forms of exploitation”. The definition is established as long as the three elements are fulfilled. Human trafficking is also known as modern slavery or trafficking in persons. These are just the umbrella terms that refer mostly to sexual exploitation and forced labor. Although the purpose of the exploitation goes beyond sex trafficking and forced labor, most of the studies related to analytics emphasized mainly on sex trafficking.

Some papers used online data from advertisements to predict the trends of sexual exploitation. One of the literature examined data from the national authorities & explored public events like Super Bowl to track sex trafficking activity [7,8,9,18]. Contrary, there are only a few papers that explored labor trafficking using statistical models [10-15]. The remaining literature studies investigated the general view of human trafficking [16-20].

The latest global report on human trafficking indicates that 50% of the exploitation was sex trafficking, 38% forced labor and 12% of other forms. The proportion of the exploitation remained almost the same over the years, which explains the immense studies in the area of sexual exploitation. Furthermore, data related to sex trafficking are more accessible from online advertisements and the national authorities because sexual exploitation is more popular and profitable for traffickers.

III. METHODOLOGY OF REVIEWING THE RELATED WORKS

Since the focus of this paper is the analytics in human trafficking, the methodology of reviewing the related studies is based on the concept and structure adopted in [5]. Words such as “human trafficking”, “human trafficking detection”, “sex trafficking detection”, “analytics in human trafficking”, “labor trafficking detection”, “modern slavery” and “data mining in human trafficking” are widely used during the literature research.

The extent of the literature review is only limited to journal articles and conference papers found in the databases of ArXiv, ScienceDirect, ACM, IEEE and SpringerLink. There are few aspects covered, as follows:

1. The literature studies are chosen in accordance to the requirements listed above;
2. The methods for the analytics in human trafficking are categorized into two parts: statistical analysis and machine learning;
3. Any combined techniques will be briefly discussed;
4. While reviewing the previous works, any identified gaps will be discussed and potential directions will be suggested; and
5. The analytics methods identified will be listed along with the purpose of exploitation for discussion.

TABLE I. THE NUMBER OF PUBLICATIONS RELATED TO ANALYTICS IN HUMAN TRAFFICKING OVER THE YEARS.

Year	No. of Papers Published
2011	1
2012	4
2013	2
2014	6
2015	4
2016	6
2017	3
2018	2
2019	2
2020	3

The earliest study that had taken advantage of data mining methods on predicting human trafficking [12]. This study had motivated more publications in the following years. The distribution can be found in Table I, where it shows that there

is a rising trend in analytics research in the area of human trafficking, especially in the year 2014 and 2016 as compared to only one related publication found in the year 2011.

IV. ANALYTICS METHODS ADOPTED BY PREVIOUS WORKS

The literature studies collected for this paper applied numerous techniques of analytics. To properly examine each technique used, this work partially adopted the classification of the methods [16]. In this paper, the methods of analytics will be categorised into two parts: statistical analysis and machine learning. The categorization of the techniques is shown in Fig. 2. Before discussing the related works, this section will explore the definition and the types of different methods in analytics.

A. Machine Learning

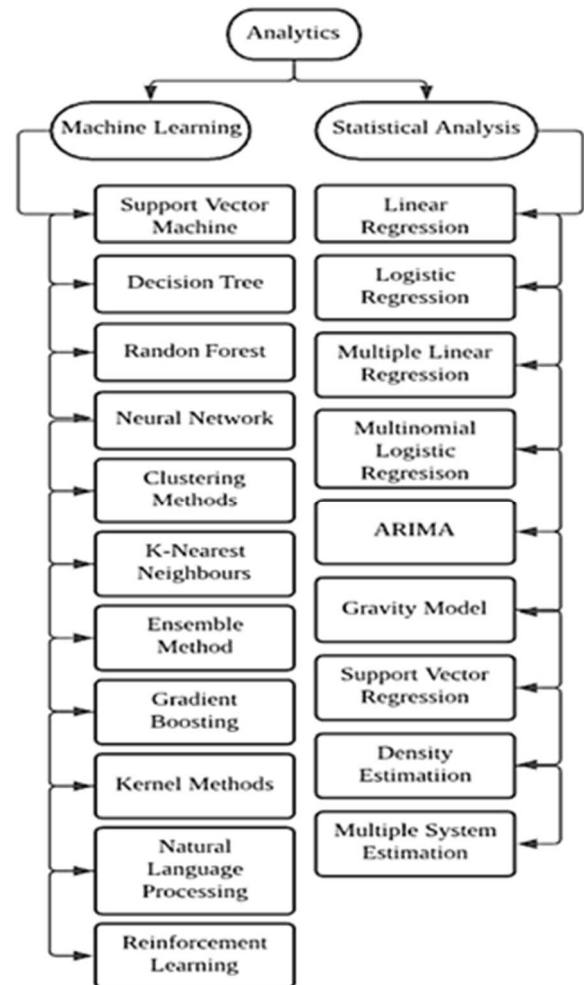


Fig. 2. The categorization of methods in analytics

Machine learning originated from the evolvement of computer strength, statistical techniques and accessible data. It basically applies computer algorithms and mathematical methods to transform data into actionable insights. It is a process that investigates the patterns within the data using statistics. Machine learning algorithms build a model from the training data (i.e., sample data) based on the instructed input to predict trends or produce insight for decision making. It can be deployed with data preparation, data pre-processing, model training, model validation and model advancement.

There are different types of machine learning algorithms, which can be categorised as supervised and unsupervised learning. Supervised learning is the process where there is a target variable, and can be used to predict an outcome (i.e., predictive model) or classify the objects (i.e., classification model). It includes, but is not limited to, models such as decision tree, random forest, support vector machine (SVM), neural network, ensemble method, gradient boosting and kernel methods. On the other hand, the process where there is no target variable for the model to learn is known as the unsupervised learning, which includes clustering and K-Nearest Neighbours (KNN). It should be noted that the list of unsupervised models in this paper is not exhaustive as shown in Fig. 2.

B. Statistical Analysis

Statistical analysis is a subset of mathematics that involves processes such as data collection, data analysis, data exploration, data interpretation and data presentation. It covers all features of data, from survey planning and collecting data to spotting patterns from the data using statistical model. The purpose of it is to explore the relationships between different features. Although machine learning algorithm and statistical analysis seem to be similar, they are significantly different from each other.

TABLE II. THE COMPILATION OF PREVIOUS WORKS BASED ON METHODS AND PURPOSE OF EXPLOITATION.

Analytics Methods	References	No.
Sexual Exploitation		
Machine learning (ML)	Keskin et al., 2020; Mensikova and Mattman, 2018; Alvari et al., 2017, 2016; Zhou et al., 2016; Miller et al., 2016; Dubrawski et al., 2015; Wang et al., 2014; Sethi et al., 2013; Latonero, 2011	10
Statistical Analysis (SA)	Maass et al., 2020; Konrad, 2019; Martin and Hill, 2019; Kosloski et al., 2017; Ibanez and Gazan, 2016a, 2016b; Roe-Sepowitz et al., 2016; Ibanez and Suthers, 2014; Dank et al., 2014; Rao and Presenti, 2012;	10
Both ML & SA	Boecking et al., 2018; Nagpal et al., 2017;	2
Forced Labor		
Machine learning (ML)	-	0
Statistical Analysis (SA)	Pitts et al., 2015; Zhang and Li, 2015; Barrick et al., 2014, 2013; Zhang et al., 2014; Zhang, 2012a, 2012b	7
Both ML & SA	-	0
All Forms of Exploitations = Human Trafficking		
Machine learning (ML)	-	0
Statistical Analysis (SA)	Sarrica, 2020; Jiang and LaFree, 2017; Cho, 2015; Hernandez and Rudolph, 2015; Silverman, 2014; Peerapeng et al., 2012	6
Both ML & SA	-	0

Machine learning applies statistical techniques and computer algorithms to mine information from the data, whereas statistical analysis utilises probability, sampling and assumptions to create inference from the data. The popular models for statistical analysis are linear regression, multiple linear regression, logistic regression, multinomial logistic regression, ARIMA, gravity model, support vector regression and density estimation.

The compilation of related works grouped by analytics methods is given in Table II. Analysis and discussion of the related works are presented in the following sections.

V. CONCLUSIONS

Human trafficking is a global, national, societal and community problem. This issue is pervasive especially during a pandemic. That being said, human trafficking still remains under-developed in the field of data science when analytics could contribute significantly. This paper reviewed the previous works related to the analytics in human trafficking. A systematic methodology was adopted so the overview of the findings could be seen clearly. Its definition was broken down so the compilation of the related works could be grouped by the purpose of exploitation, referenced studies and the number of published works.

Surprisingly, the analysis of the related works revealed that there are still a number of areas that remain untouched, especially the area of forced labor. There were no machine learning algorithms deployed to detect or predict forced labor. It has been suggested that there is limited data in the field of human trafficking, which made it difficult to apply analytics, however, creativity is seen from the previous works that have leveraged on any available open-source data to tackle this issue. Therefore, leveraging on any accessible data with care and innovative approach should be adopted for future research.

REFERENCES

- [1] Alvari, H., Shakarian P. and Snyder, J. (2016) A non-parametric learning approach to identify online human trafficking. In IEEE International Conference on Intelligence and Security Informatics: Cybersecurity and Big Data. Institute of Electrical and Electronics Engineers. p. 133–138.
- [2] Alvari, H., Shakarian, P. and Snyder, J. (2017) Semi-supervised learning for detecting human trafficking. *Security Informatics*. 6 (1). p. 1-14.
- [3] Barrick, K., Lattimore, P. K., Pitts, W. J. and Zhang, S. X. (2014) What law enforcement agencies fail to see, what farmworkers and advocacy groups see: Challenges in identifying labor trafficking in North Carolina. *Crime, Law and Social Change*. 61. p. 205–214.
- [4] Barrick, K., Lattimore, P. K., Pitts, W. J., & Zhang, S. X. (2013) Indicators of labor trafficking among North Carolina migrant farmworkers (Final report for NIJ Grant Number 2009-IJ-CX-0047). Washington, DC: National Institute of Justice.
- [5] Boecking, B., K. Miller, E. Kennedy, A. Dubrawski (2018) Quantifying the Relationship between Large Public Events and Escort Advertising Behavior. *Journal of Human Trafficking*. 5 (3). p. 220-237.
- [6] Boyd, A. (2012) Revisiting 'what is analytics'. *Analytics Inform.* (August), Available from: <https://pubsonline.informs.org/doi/10.1287/LYTX.2012.04.09/abs/>. [Accessed: 20 March 2021].
- [7] Cho, S. Y. (2015) Modeling for determinants of human trafficking: An empirical analysis. *Social Inclusion*. 3. p. 2–21.
- [8] Dank, M., Khan, B., Mitchell Downey, P., Kotonias, C., Mayer, D., Owens, C., Pacifici, L., & Yu, L. (2014) Estimating the size and structure of the underground commercial sex economy in eight major US cities. Research Report March 2014, The Urban Institute. p 16–20.

- [9] Dubrawski, A., Miller, K., Barnes, M., Boecking, B. and Kennedy, E. (2015) Leveraging publicly available data to discern patterns of human-trafficking activity. *Journal of Human Trafficking*. 1. p. 65–85.
- [10] Hao, K. (2018) What is machine learning? MIT Technology Review. (November), Available from: <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/> [Accessed: 9 March 2021].
- [11] Hernandez, D., and Rudolph, A. (2015) Modern day slavery: What drives human trafficking in Europe? *European Journal of Political Economy*. 38. p. 118–139.
- [12] Ibanez, M. and Gazan, R. (2016a) Detecting sex trafficking circuits in the U.S. through analysis of online escort advertisements. In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE. p. 892–895.
- [13] Ibanez, M. and Gazan, R. (2016b) Virtual indicators of sex trafficking to identify potential victims in online advertisements. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). San Francisco, USA: CA. p. 819–824.
- [14] Ibanez, M. and Suthers, D. D. (2014) Detection of domestic human trafficking indicators and movement trends using content available on open internet sources. In 2014 47th Hawaii International Conference on System Sciences. p. 1556–1565.
- [15] International Labour Office and Walk Free Foundation (2017) Global estimates of modern slavery: forced labour and forced marriage. [Online]. Available from: https://www.ilo.org/wcmsp5/groups/public/@dgreports/@dcomm/documents/publication/wcms_575479.pdf [Accessed: 19 March 2021].
- [16] Jiang, B., and LaFree, G. (2017). Social control, trade openness and human trafficking. *Journal of Quantitative Criminology*. 33(4). p. 887–913.
- [17] Keskin, B. B., Bott, G. J. and Freeman, N. K. (2020) Cracking Sex Trafficking: Data Analysis, Pattern Recognition, and Path Prediction. *Production and Operations Management*. POMS 13294. p. 1–40. DOI: 10.1111/poms.13294.
- [18] Konrad, R. A. (2019) Designing awareness campaigns to counter human trafficking: An analytic approach. *Socio-Economic Planning Sciences*. 67. p. 86–93.
- [19] Kosloski, A., Bontrager R. E. S. and Roe-Sepowitz, D. (2017) Buying the Girl Next Door: A Study of Solicitations for Sex Online. *Family & Intimate Partner Violence Quarterly*. 9 (4). p. 1–12.
- [20] Lantz, B. (2019) *Learning with R: Expert techniques for predictive modelling*. 3rd ed. Birmingham: Packt.