Annals of Operations Research (2024) 335:761–795
https://doi.org/10.1007/s10479-023-05520-1

ORIGINAL RESEARCH

Disentangling human trafficking types and the identification of pathways to forced labor and sex: an explainable analytics approach

Enes Eryarsoy[1] · Kazim Topuz[2] · Cenk Demiroglu[3]

Received: 24 August 2022 / Accepted: 7 July 2023 / Published online: 28 July 2023

Abstract

Terms such as human trafficking and modern-day slavery are ephemeral but reflect manifestations of oppression, servitude, and captivity that perpetually have threatened the basic right of all humans. Operations research and analytical tools offering practical wisdom have paid scant attention to this overarching problem. Motivated by this lacuna, this study considers two of the most prevalent categories of human trafficking: forced labor and forced sex. Using one of the largest available datasets due to Counter-Trafficking Data Collective (CTDC), we examine patterns related to forced sex and forced labor. Our study uses a two-phase approach focusing on explainability: Phase 1 involves logistic regression (LR) segueing to association rules analysis and Phase 2 employs Bayesian Belief Networks (BBNs) to uncover intricate pathways leading to human trafficking. This combined approach provides a comprehensive understanding of the factors contributing to human trafficking, effectively addressing the limitations of conventional methods. We confirm and challenge some of the key findings in the extant literature and call for better prevention strategies. Our study goes beyond the pretext of analytics usage by prescribing how to incorporate our results in combating human trafficking.

Keywords Human trafficking · Forced sex · Forced labor · Machine learning · Bayesian belief networks · Analytics

✉ Enes Eryarsoy
enes@sabanciuniv.edu

Kazim Topuz
kazim-topuz@utulsa.edu

Cenk Demiroglu
cenk.demiroglu@ozyegin.edu.tr

1 Sabanci Business School, Sabanci University, Istanbul, Turkey

2 Collins College of Business, School of Finance and Operations Management, The University of Tulsa, Tulsa, OK 74104, USA

3 Electrical and Electronics Engineering, Faculty of Engineering, Ozyegin University, Istanbul, Turkey

1 Introduction

The time of penning this manuscript coincided with the observation of World Day Against Trafficking in Persons, July 30. While the day was proclaimed by the United Nations (UN) as early as 2014, and the protocol to prevent, suppress, and punish was adopted in 2018 (United Nations, 2018), the issue of human trafficking (HT) is more pressing than ever. Data on HT is limited, yet it is not difficult to cite a startling HT-related statistic that will appeal to any audience. The UN Office on Drugs and Crime (UNODC) estimates over 25 million HT victims globally, yet only around 20 K are convicted annually (UNODC, 2020) (Fig. 1). According to an article in Huffington Post, from coffee beans and footwear to the tantalum used in their smart phones, the typical hipster's purchasing habits employ as many as 27 slaves (Couch, 2015). In another article in NYT Paton and Ramzy (2020) mention that one in every five cotton garments globally has cotton from Xinjiang, a region with government-imposed labor. The UN (2015) lists the implementation of "rapid and effective steps to eliminate modern slavery and HT" as an objective under its Sustainable Development Goal of "decent work and economic growth". The U.S. Department of Justice recently awarded $101 million to battle HT in 2020 (The U.S. Department of Justice, 2020). However, according to the International Labour Organization (ILO), forced labor (FL) alone produces an estimated $150 billion in profits per annum (ILO, 2014). When added up together, these figures show that HT organizations fighting this profitable form of human rights violation do exist but are outnumbered.

It is not straightforward to measure the reach of HT. There are estimates of its magnitude. However, some resources cite conflicting figures. For example, Whitman and Gray (2015) mention that around 600–800 thousand women are trafficked each year. This number is substantially lower than that reported by the UNODC, which estimates around 25 million HT cases and identifies sexual exploitation as the most prevalent form of HT among the

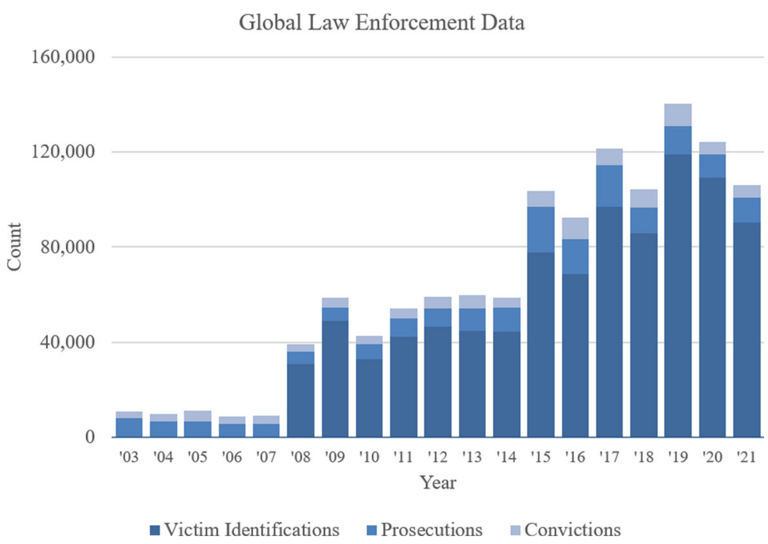


Fig. 1 Visualization of global law enforcement data on HT. Combined from 2022, and 2018 "Trafficking in Persons Reports" (U.S. Department of State, 2018; U.S. Department of State, 2022a, 2022b). While progress has been made, the numbers are low compared to 35 M HT victims

detected victims (UNDOC, 2020). According to ILO, of the 50 million living in modern slavery as of 2021, 27 million are in forced labor (FL), including forced sex (FS). Their figures also suggest FL and FS are the most prevalent form of modern slavery, however, differ in that FS accounts for less than 30% (ILO, 2022). According to an ILO report (ILO, 2014) based on interviews with over 71,000 respondents from 48 countries, there are 40 million trafficked people at any given time. Modern slavery, which is distinct from human trafficking but closely interrelated, represents yet another pervasive and paramount issue. The aforementioned report also reveals that over the last 5 years, almost 90 million people have been subjected to some form of modern slavery, with 25% being children and three-quarters being women, and debt bondage being the most frequent mean of control (50%). Besides debt bondage, other means of control for modern slavery, including FL, involve debt bondage, descent-based slavery via maternal line, child slavery, and forced and early marriage (ILO, 2014). The report also notes "Due to limitations of the data, these estimates are considered to be conservative". Authorities also rank HT as the second most lucrative and fastest-expanding illicit activity after drug trafficking (Sheinis, 2012). They project that HT will surpass drug and arms trafficking in frequency, damage to human well-being, and criminal profitability (Schauer & Wheaton, 2006). With significantly smaller conviction rates, HT is paving its road to overtake the leadership position. Although these estimates vary, they all point in the same direction.

Thanks to their stipulatory definitions, the crime of selling drugs and arms illegally is more measurable and easier to combat than HT (UNSC, 2017). Regarding HT, Wheaton et al., (2010, p. 133) state "A felony lacking definition cannot be controlled, nor can it be counted or measured". This is one of the many factors that make combating HT more difficult. Because some labels, including HT, can be regarded as social constructions (Hoyle et al., 2011). Often the label is decided not only by those who were hurt but also by the views of those who may attach or reject such labels. Different regulatory authorities can and do classify HT victims with differing rationales. A universal identification of the crime is merely one of the many obstacles encountered in the fight. For example, the UK Visas on Immigration is less eager than the UK National Crime Agency to identify individuals as HT victims. Analogously, the U.S. Immigration and Customs Enforcement (ICE) and the Department of Health and Human Services (HHS) might classify HT victims differently; the former is under the Department of Homeland Security and is concerned with apprehending traffickers and their victims' eligibility for immigrant visas, while the latter is focused on assisting HT victims regardless of their immigration status.

To counteract and prevent HT, it is necessary to better understand the interaction between different forms of HT and the means of control employed by offenders. The U.S. Department of State employs the 3P paradigm, prosecution of traffickers, protection of victims, and prevention of HT, equivalent to the UN's prevent/suppress/punish protocol, as the cornerstone for combating HT worldwide (The U.S. Department of State, 2022a, 2022b). While the basic goal of HT is to generate a monetary return (UNODC, 2020), different forms of HT are influenced by various factors and actors. These influencers mostly determine the level of difficulty associated with fighting the crime. For example, fighting a state-imposed HT-driven FL, such as in Xinjiang, may demand a more coordinated approach. According to the Counter-Trafficking Data Collective (CTDC), in many HT cases involving sex and children, the victims are purposefully exploited by their family members in exchange for drugs or money (CTDC, 2022). It is important to note that although FL and FS are strongly linked to HT, the occurrence of FL and FS does not necessarily hinge on the existence of HT. HT is also distinct from FS and FL, despite being closely interrelated. While HT involves the transportation of individuals from one location to another, it does not necessarily involve their exploitation primarily for economic activities (Crane, 2013). Understanding such FL and FS paths may assist policymakers in developing a more coordinated approach. Different means, such as threats or physical abuse, used by certain actors, such as a family member or an intimate partner, may exacerbate a potential victim's predisposition to a specific form of HT. This study aims to address and narrow this gap by analyzing the first known global dataset on HT maintained by CTDC (CTDC, 2022).

Our manuscript is structured as follows. In Sect. 2, we provide a literature review. Following the review, we introduce the problem definition and our framework. We then introduce the dataset and present the findings of our descriptive analysis. In Sect. 4, we undertake two sets of analysis: (i) we build a logistic regression (LR) model to characterize different kinds of HT, (ii) we also create a Bayesian Belief Network (BBN) based decision support tool to help decision-makers understand the interplay among the available variables. We discuss our findings and surrounding assumptions in Sect. 5. The final section constitutes the conclusion and future research.

2 Literature review

In their 2013 publication, Crane (2013) investigates the underlying factors that enable various types of modern slavery to persist and proposes a theoretical model that encapsulates these factors. Their model identifies the "enabling conditions" and "exploiting capabilities" that contribute to modern slavery. While FL and FS do not necessarily hinge on the existence of HT, they are inextricably linked. There is no "one-stop shop" for solution recipes when combating a multibillion-dollar enterprise such as HT. Even though the literature on HT is fast expanding to match the need for solutions, there seems to be a detectable level of disagreement and variation across studies (Okech et al., 2018). This variation is detectable as early as the definition and taxonomic identification of different HT forms. Different researchers/organizations offer varying definitions and taxonomies (E.g., seven HT categories due to Hachey and Phillippi (2017) vs. ILO's (2014) three-level taxonomy). Nonetheless, everyone acknowledges FL and FS as the two largest forms of exploitation associated with human trafficking (HT) and modern slavery. The literature also embodies alternative conceptual models to understand HT. Caltagirone (2017) presents a seven-step linear cognitive model for the HT chain recruitment, transportation, entrapment, brokering, delivery, and exploitation. The following phases constitute their plan of action against this chain (6Ds): detect, deny, disrupt, degrade, deceive, and destroy. The 4P (prevent protect prosecute and partnership) is an alternate framework for a course of action (Dimas et al., 2020; Office to Monitor and Combat Trafficking in Persons, 2019). These demonstrate that HT literature is still developing (Sweileh, 2018).

Several studies reviewed HT literature. However, to the best of our knowledge, four recent dedicated studies have sought to review the literature on HT. Sweileh (2018) conducted a bibliometric analysis of the Scopus database from 2000 to 2017, analyzing keywords, authorship, citation, and international cooperation. Their analysis grouped HT literature across three dimensions: (i) growth (ii) health vs. non-health-related aspects, and (iii) source vs. destination countries. Okech et al. (2018), in their systematic literature review, found out that the bulk of HT research effort went into sex trafficking and emphasized the absence of a precise characterization for HT and its variants. Cockbain et al. (2018) focused on HT and FL in their two-phased systematic literature review. They reported that the qualitative designs outweighed the quantitative ones in the extant literature and that many of the research conclusions were not properly grounded in the data. Perhaps the only study that reviewed and classified the quantitative designs is by Dimas et al. (2020). In their outstanding research, Dimas et al. (2020) reviewed and classified 142 studies in the extant HR-literature through OR and analytics lenses. They categorized the literature according to the “4Ps course of action” (partnership, in addition to the 3Ps employed by the U.S. Department of State: prevention, protection, and prosecution) and according to the OR and analytic techniques employed in the study. They discovered that nearly half of the research focused on inferential statistics. We recommend that readers read the publication by Dimas et al. (2020) for an outstanding discussion of a broader range of OR and analytics applications on HT. The following paragraphs are dedicated to more relevant works linked to our research.

While OR and analytics fill in critical gaps in the fight against HT, the bottleneck in the analytical front is mostly the lack of data. In their study, Dimas et al. (2022) provided a very good overview of how OR and analytics could be used in combating HT. In terms of a unit of analysis, a small percentage of the extant literature analyzes HT at the country-level. Bales (2007) conducted a country-level analysis to explore the HT linkage between countries. Ramchandani et al. (2021) attempted to paint a network picture of HT supply chains. Case-level analyses seem to me much more common in analytical research. Several studies have employed information visualization tools and social media data to understand and identify HT cases (Alvari et al., 2016, 2017; Andrews et al., 2018). Even though they are needed, these efforts still cannot make up for the lack of data. Alvari et al. (2017) used semi-supervised learning techniques to boost dataset size for their analysis.

One of the reasons for the shortage of case-level data is its sensitivity. Recently, CTDC and Microsoft collaborated to create a synthetic HT dataset through anonymization that conceals sensitive case-level details. A few recent studies have used the CTDC dataset. The majority of them have conducted descriptive analyses of the dataset. Szakonyi et al. (2021) utilized the random forest technique to predict the victim’s "Gender" and ranked the relative importance of the variables. They discovered that the country of exploitation, citizenship, and type of exploitation are the most significant predictors. A bulk of HT studies made use of datasets collected from online advertisements (e.g. Chen et al., 2015; Alvari et al., 2016; Dubrawski et al., 2015; Giommoni & Ikwu, 2021; Kejriwal & Szekely, 2017a, 2022; Keskin et al., 2021; Ramchandani et al., 2021), and from crowd-sourced review platforms (Diaz & Panangadan, 2020) and social media (Andrews et al., 2018; Burbano and Hernandez-Alvarez, 2018a, 2018b). While less common, some studies used data collected from confirmed HT cases. Cockbain et al. (2011) used data from 23 offenders and 36 victims and suggested that social network analysis could be useful. Poelmans et al. (2012) created early warning indicators to index the police reports using textual analysis and concept lattice visualization tool based on 266 K suspicious activity reports in Amsterdam. They mention that, for the cases they uncovered, new evidence verified the participation of the exposed suspects in HT. In another study, da Silva Santos et al., (2019a, 2019b) confirmed suspected HT cases using data from the Brazilian Secretary of Labor. Kiss et al. (2021) used an extracted dataset on 519 Nepalese refugees to design a Bayesian network mapping different paths leading to HT. Libaque-Saenz et al.(2017) used Peruvian National Housing Survey to predict child labor. Some optimization-based analytic work has also been done in the fight against HT. Konrad (2019) developed a resource allocation model with three potential objectives, including the maximization of the level of awareness and providing resources proportionally to HT rates of incidence. Maass et al. (2020) proposed an optimization model for shelter location for HT survivors. Dimas et al. (2022) estimated the effectiveness of combating HT using data envelopment analysis.

Our work shows some resemblance to these two studies: Cockbain and Bowers (2019) and Poelmans et al. (2012). Similar to Cockbain and Bowers (2019) we also employ predictive analytics in tackling this problem. However, our study differs in two particular ways. First, instead of employing logistic regression, we employ Bayesian belief networks that fit the data better (a much higher accuracy than what was reported in their study: 76%), that provide greater transparency (it is based on conditional probability tables), and can be used as a decision tool (WebSimulator). Second, we differ in that we intentionally omit some key variables such as citizenship (77 different countries exist in the dataset), in our analysis. Some variables, such as "Gender" have more predictive power than others. For instance, Cockbain and Bowers (2019) report gender as "the most pronounced" and note "being female means someone is 75 times more likely to have been trafficked for sex than labor". While we do not omit the "Gender" variable, WebSimulator does show the predictive power of "Gender". We acknowledge and illustrate these variables' predictive power in our study. However, our dataset contains a variety of additional variables, allowing a more in-depth investigation and a richer discussion in the absence of these omitted variables. In the second study, Poelmans et al. (2012) developed a concept hierarchy for the available cases using formal concept analysis. Their use of FCA for knowledge discovery from short unstructured texts is similar to our use of BBNs. FCA is the predecessor of association rules analysis (Ganter et al., 2005). FCA's focus is identifying frequent co-occurrence of items, while BBNs model probabilistic dependencies between variables. In this study, we focus on the creation of a probabilistic decision tool.

3 Problem description and research method

3.1 Research methodology

While the availability of public and sanitized datasets in fighting HT is precious (Cockbain et al., 2018), they come with certain considerations. First and foremost, the shortcomings of the dataset (in terms of size, accuracy, and bias) cannot nourish all types of data-driven research. The shortcomings guide the type of research that could be conducted. Given the dataset's limitation, the objective of this study is to scrutinize two primary forms of HT in the CTDC dataset: FS and FL. According to ILO (2014), FS is a type of FL, and FL and forced marriage are two of the major categories in HT. However, the CTDC dataset covers only a few forced marriage cases (1.7% of the records). Its records primarily correspond to the FS and FL cases (67 and 29% respectively, accounting for 96% of the dataset after discounting the records with missing values in the "Abuse Type" variable). We limit our study to FS and FL cases. This is reasonable, as authorities identify both FT and FS as the most prevalent forms of modern slavery (ILO, 2022; Cockbain & Bowers, 2019; Feehs & Currier, 2020; UNODC, 2020). ILO (2022) indicates that 23% relates to FS and 63% to private sectors other than commercial sex, according to a report released in collaboration with the WalkFree foundation and the UN migration. Some sources (Cockbain & Bowers, 2019; Human Trafficking Institute, 2021) report higher figures for FS. UNODC (2020) data reports similar figures for FS and FL rates to the dataset as 50% and 38%, respectively. The difference can be attributed to reporting bias. According to Farrell et al., (2020), U.S. police efforts are more focused on sex trafficking than labor. Cockbain and Bowers (2019, p. 10) state that "responses to trafficking for purposes other than sex remain comparatively underdeveloped". In our analysis, we recognize that the true distribution of abuse type may deviate from the observed frequencies in our dataset.

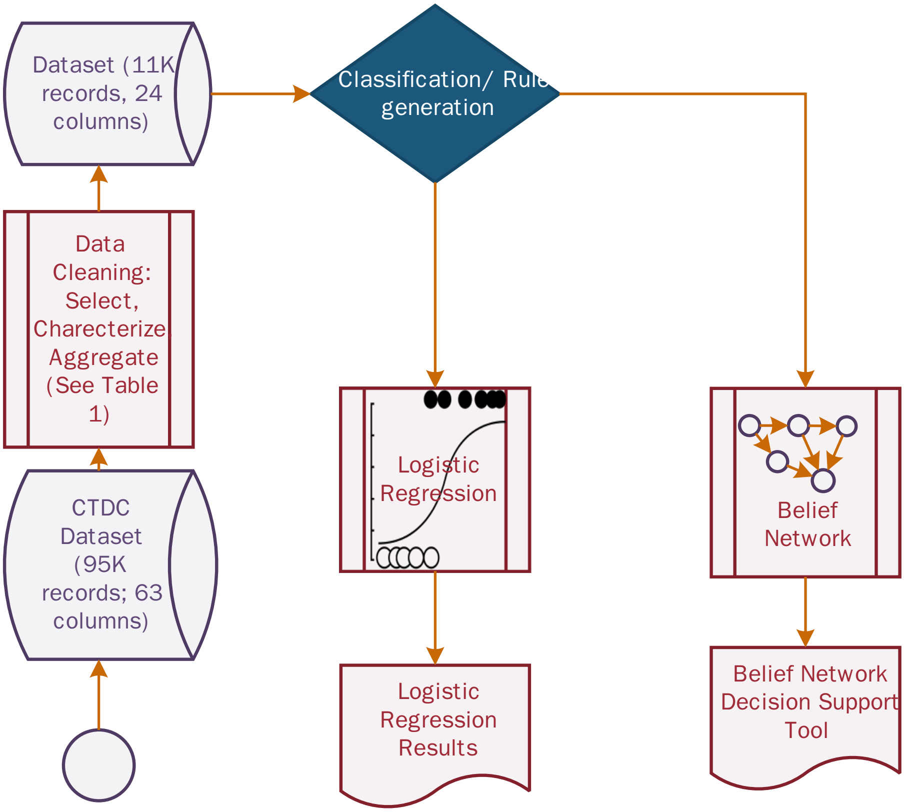


Fig. 2 Graphical representation of our approach

To lay the foundation of our work, we present the key steps of our analysis in this section. Figure 2 shows the graphical representation of our procedure.

3.2 Data acquisition and preparation

As noted, data scarcity is still one of the most pressing issues in the discovery and generation of actionable insights for law enforcement and policymakers, notwithstanding the significance of the HT problem (Dimas et al., 2020; ILO, 2014; Alvari et al., 2017). The available datasets are prone to biases such as coverage bias (i.e. the majority of cases are from the U.S.), and sampling bias (i.e. there are more FS instances in our dataset, but reliable sources estimate a greater prevalence for FL). *CTDC, a global data hub founded in 2017, is dedicated to collecting and distributing data from anti-HT organizations.* To the best of our knowledge, the CTDC dataset is the largest among the small number of publicly available ones (CTDC, 2022). The dataset is collected at the case-level, namely, it is organized and presented at the level of individual cases or incidents. Here, case-level refers to the granularity and the organization of the dataset. While some of these might be associated with legal proceedings, the dataset does not include such information. The CTDC dataset combines data mainly from two parties: International Organization for Migration (IOM) and the non-profit organization Polaris which combats HT. While Polaris data comes from U.S. National Human Trafficking Hotline and the BeFree Textline, IOM collects data on trafficking through its counter-trafficking programs.

However, the aforementioned biases exist in the CTDC dataset as well. The dataset is comprehensive, but not a random sample of the HT-victim population. For instance, upon request, Polaris removes information about individuals who do not wish to be included in the dataset. The dataset also doesn't specify the origin of each record. It, however, indicates whether the record is collected through "hotline" (2015–2019) or "case management" (by IOM, 2002–2021). While such biases render the representativeness of the sample problematic, it is still one of the finest sources out of the few that are accessible. To protect the anonymity of the victim, the dataset is k-anonymized (CTDC, 2020; Sweeney, 2002). The codebook and data dictionary are made available on the CTDC's website (CTDC, 2021). The original dataset contains 63 variables, which we broadly divide into four categories: (i) variables related to demographics and record-specific key information such as data source, citizenship, and year (ii) variables capturing the means of control, (iii) variables associated to HT type, (iv) variables representing the link between HT-victim and its recruiter. In its original form, with 95,739 records, the dataset was unsuitable for our investigation. The dataset was not suitable for our analysis. Therefore, aside from cleaning up and fixing a few small mistakes, we follow the procedures outlined in Table 1 to extract our analysis dataset. We also do not admit the "Citizenship" variable in our analysis. We evaluate and confirm its predictive value but intentionally exclude the citizenship variable for two reasons: to avoid excessive accuracy and bias and for practical reasons such as being able to put results into action through the

Table 1 Data cleaning steps

Step	Description
Step 1	Removal of the rows: (i) without any "Means of control" specified (51,287 records) (ii) without any "victim-trafficker relationship" specified (16,549 records)
Step 2	Removal of the rows that are related to other types of exploit such as "Forced marriage" (3,738 records), "abduction" (264 records), "forced military" (3862 records), "organ removal" (3862), "slavery and practices" (26 records), and other unspecified exploit types- (868 records)
Step 3	Removal of the records without either FL or FS (5926 records)
Step 4	Creation of "Minority status change" variable: During the case-collection process, some minors who were minors at the time of HT. We, therefore, created a new variable that identifies victims that were minors at the time of the exploit but became adults before the case was reported and removed the columns indicating the minority status of the victims. (1922 of the filtered 11,368 records)
Step 5	Removal of the "Citizenship" variable: This is a variable with high predictive power (See Sect. 3.2 for details). The original collection included 61 nations, with a large fraction (> 25 percent) of the vales missing. The predictive power of citizenship, within the scope of our study, is not relevant
Step 6	Removal of irrelevant columns. The original dataset included irrelevant or redundant columns such as three concatenated variables. We first cross-check and validate the concatenation to assure data integrity. Then, following the study objectives, we eliminate the redundant columns
Step 7	Removal of country of exploitation: While a total of 77 countries are mentioned, most records (> 70%) correspond to the U.S
Step 8	Fixing missing values: We interpret missing values associated with "means of control" or "relationship" as absence (zero rather than missing) and check for inconsistencies across all related columns

Table 2 Variable coding and descriptive statistics for logistic regression ($n = 11,368$)

	Variable	Percentage		Variable	Percentage
		62	Means of control	Threats	42.1
		38		Restricts movement	36.4
	FS	69.9		Takes earnings	33.9
	FL	30.1		False promises	3.0
				Physical abuse	29.8
	Recruiter relation			Excessive working hours	27.1
Recruiter Relation	Intimate partner	28.6		Withholds documents	2.9
	Other	19.1		Withholds necessities	17.9
	Family	19.1		Threat of law enforcement	17.6
	Friend	12.1		Psychoactive substances	16.4
				Sexual abuse	13.7
				Debt bondage	13.0
				Restricts medical care	12.5
				Minority status change	6.0
				Restricts financial access	2.1
				Uses children	1.8
				Psychological abuse	69.8
				Other	42.3

creation of awareness campaigns where citizenship cannot be determined. The resulting final dataset has 11, 368 rows and 24 columns. Tables 2 and 3 provide descriptive statistics and correlation tables, respectively. All the correlation coefficients' absolute values were less than 0.75.

More than 90% of HT-related cases in the United States involve either FL or FS (FS being the more prevalent form of abuse). This is congruent with the UNODC study (UNODC, 2020), which identifies sexual exploitation (79%) as the most prevalent form of HT; and with the National Crime Agency (2018), which attributes 34% of suspected victims to modern slavery, 46% to labor exploitation and 9% to domestic servitude.

3.3 Logistic regression

We begin with visualizing our dataset to gain insights. The dataset includes information about the two types of abuse, FL and FS, and two main categories of variables: relationship to exploiter, and means of control (see Appendix A for their explanations). Figure 3 links

Table 3 Correlation table for the dataset ($n = 11,368$)

	[V1]	[V2]	[V3]	[V4]	[V5]	[V6]	[V7]	[V8]	[V9]	[V10]	[V11]	[V12]
Gender [1]	–											
Minority Status Change [2]	−0.14	–										
Debt Bondage [3]	0.17	−0.07	–									
Takes Earnings [4]	0.44	−0.12	0.27	–								
Restricts Financial Access [5]	−0.02	−0.02	0.02	0.04	–							
Threats [6]	0.20	−0.06	0.18	0.33	0.03	–						
Psychological Abuse [7]	0.00	0.08	0.01	0.02	−0.03	0.03	–					
Physical Abuse [8]	−0.11	0.02	0.05	0.10	0.02	0.19	0.11	–				
Sexual Abuse [9]	−0.21	0.12	−0.07	−0.14	0.03	0.00	0.01	0.12	–			
False Promises [10]	0.40	−0.13	0.27	0.49	0.00	0.25	−0.03	0.02	−0.17	–		
Psychoactive Substances [11]	−0.25	0.05	−0.10	−0.17	−0.02	−0.02	−0.04	0.10	0.17	−0.22	–	
Restricts Movement [12]	0.22	−0.08	0.21	0.35	0.08	0.33	0.04	0.20	−0.03	0.35	−0.01	–
Restricts Medical Care [13]	0.26	−0.07	0.26	0.41	0.01	0.29	0.11	0.23	−0.11	0.46	−0.10	0.33
Excessive Working Hours [14]	0.60	−0.14	0.23	0.58	−0.03	0.30	0.08	0.02	−0.21	0.56	−0.25	0.38
Uses Children [15]	−0.06	0.01	−0.01	−0.03	0.05	0.04	0.06	0.07	0.04	−0.06	0.05	0.00
Threat Of Law Enforcement [16]	0.47	−0.11	0.16	0.31	−0.03	0.25	−0.01	−0.02	−0.15	0.23	−0.16	0.14
Withholds Necessities [17]	0.23	−0.07	0.11	0.31	0.08	0.17	−0.12	0.09	−0.04	0.19	−0.08	0.18
Withholds Documents [18]	0.38	−0.12	0.27	0.50	0.02	0.32	0.13	0.13	−0.14	0.54	−0.16	0.47
Other [19]	0.02	−0.01	0.00	−0.03	0.04	0.05	−0.02	0.04	0.08	−0.09	0.05	0.03
Intimate Partner [20]	−0.41	−0.01	−0.16	−0.27	0.03	−0.11	0.28	0.16	0.06	−0.37	0.18	−0.12

Table 3 (continued)

	[V1]	[V2]	[V3]	[V4]	[V5]	[V6]	[V7]	[V8]	[V9]	[V10]	[V11]	[V12]
Friend [21]	0.11	−0.06	0.02	0.11	0.00	0.05	−0.14	0.01	0.02	0.15	−0.02	0.04
Family [22]	−0.18	0.29	−0.15	−0.27	−0.04	−0.22	0.19	−0.08	0.14	−0.25	0.02	−0.22
Other [23]	0.43	−0.16	0.25	0.38	0.00	0.27	−0.28	−0.06	−0.15	0.39	−0.11	0.27
Type of Abuse (FS = 1) [24]	0.73	−0.18	0.26	0.59	0.02	0.29	−0.06	−0.03	−0.25	0.55	−0.32	0.33
	[V13]	[V14]	[V15]	[V16]	[V17]	[V18]	[V19]	[V20]	[V21]	[V22]	[V23]	[V24]
Gender [1]												
Minority Status Change [2]												
Debt Bondage [3]												
Takes Earnings [4]												
Restricts Financial Access [5]												
Threats [6]												
Psychological Abuse [7]												
Physical Abuse [8]												
Sexual Abuse [9]												
False Promises [10]												
Psychoactive Substances [11]												
Restricts Movement [12]												
Restricts Medical Care [13]	–											
Excessive Working Hours [14]	0.49	–										
Uses Children [15]	−0.03	−0.05	–									
Threat Of Law Enforcement [16]	0.30	0.45	−0.03	–								
Withholds Necessities [17]	0.31	0.31	−0.01	0.26	–							

Table 3 (continued)

	[V13]	[V14]	[V15]	[V16]	[V17]	[V18]	[V19]	[V20]	[V21]	[V22]	[V23]	[V24]
Withholds Documents [18]	0.47	0.58	−0.04	0.23	0.16	–						
Other [19]	−0.02	−0.01	0.05	0.14	0.05	−0.10	–					
Intimate Partner [20]	−0.21	−0.38	0.11	−0.26	−0.17	−0.26	0.06	–				
Friend [21]	0.05	0.06	−0.03	0.00	0.08	0.05	−0.03	−0.21	–			
Family [22]	−0.15	−0.25	0.01	−0.21	−0.15	−0.23	−0.12	−0.27	−0.16	–		
Other [23]	0.27	0.49	−0.08	0.40	0.22	0.37	0.10	−0.51	−0.28	−0.39	–	
Type of Abuse (FS = 1) [24]	0.42	0.76	−0.06	0.52	0.34	0.54	0.02	−0.48	0.10	−0.31	0.60	–

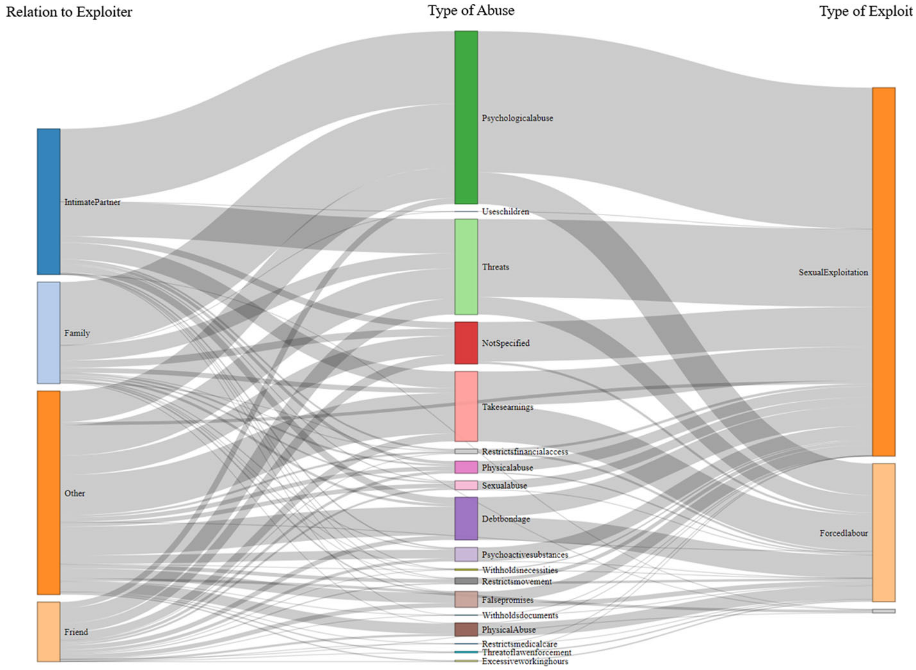


Fig. 3 Sankey diagram linking the type of exploitation to the relationship to the exploiter, and the type of abuse

these two sets of variables to the types of abuse. The figure shows, for instance, offenders who are family members or intimate partners are likelier to abuse their victims psychologically. Additionally, debt bondage is more prevalent in FL whereas threats are more typical in sexual exploitation. Although the plot offers some useful insight, it does not provide much in terms of how the means of control interact with relationship or gender regarding the type of exploit. Therefore, further analysis is needed. A broad range of techniques, particularly those that work well with dichotomous variables, such as logistic regression, decision trees, or random forests could be used. We choose logistic regression due to its simplicity and transparency, and the fact that it was used in earlier studies. To better understand the interplay between the type of abuse and the factors about gender, the trafficker's connection to the HT victim, and the means of control, we construct a regression model.

Ordinary linear regression (OLS) models with multiple independent variables have been proven to be very powerful and transparent analytical tools. However, the conventional OLS often produces subpar results when the dependent variable Y is binary. For binary dependent variables, the predicted values are likelier to be either above one or below zero. Additionally, in such cases, the OLS model won't adhere to homoskedasticity and normality assumptions. Logistic regression (LR) with maximum likelihood estimation addresses these shortcomings. In LR, the probability π corresponds to the event probability $P(Y = 1)$. Writing π with exponentials produces values between 0 and $+\infty$. This bounds π to a value between 0 and 1 as in Eq. (1). The ratio $\frac{\pi}{1-\pi}$ is called the odds ratio. The odds can assume any value in $(-\infty, +\infty)$. By taking the natural logarithm of these odds, logistic regression can be written analogously to OLS as in Eq. (2).

$$P(Y=1)=\pi=\frac{e^{\alpha+\sum_k^n \beta_k x_k}}{1+e^{\alpha+\sum_k^n \beta_k x_k}} \tag{1}$$

$$\log\left(\frac{\pi}{1-\pi}\right)=\alpha+\beta_1 x_1+\cdots+\beta_n x_n \tag{2}$$

3.4 Association rules analysis (ARA)

ARA has been successfully used to identify interesting patterns in the form of probabilistic "if–then-else" statements. These patterns denoted as $A \rightarrow B$, are called rules. Each rule is a conditional structure that identifies an existing relationship between two events, a set of antecedent events (A) and a consequent event (B). The antecedent part typically includes at least one or many events/items while the consequent (B) includes only one. The ARA engine extracts rules in two steps: frequent itemset generation, followed by rule extraction. For larger binary datasets the evaluation of every possible itemset using a brute-force approach is computationally infeasible. Therefore, different algorithms are used to construct frequent itemsets, based on user-specified thresholds.

After the extraction of frequent itemsets, the final set of "strong" rules in the form of "$A(Antecendent) \rightarrow B(Consequent)$" is created by filtering the generated rules based on specified measures of strength. Confidence, support, and lift are the three most preferred measures of rule strength used in ARA. While support refers to the frequency of the itemset $(A \cup B)$, confidence refers to how frequently the itemset $A \cup B$ appears among all records that contain the antecedent itemset A. Lift, on the other hand, is the ratio of the confidence to the expected confidence for each rule ($\mathrm{Lift}(A \rightarrow B)=\frac{\mathrm{Support}(A\cup B)}{\mathrm{Support}(A)\times \mathrm{Support}(B)}$). In other words, support, confidence, and lift are used to measure the amount (of evidence), confidence (about the evidence), and (the lack of) coincidence for each rule respectively.

3.5 Bayesian belief networks

Bayesian belief networks (BBNs) are probabilistic graphical models that illustrate the relationships between variables. These networks enable reasoning with uncertain or incomplete data, making them useful for prediction, classification, and decision-making challenges. Using Bayesian probability theory as the underlying mathematical framework, Judea Pearl presented BBNs in the 1980s as a mechanism for capturing and controlling uncertainty in probabilistic reasoning. A BBN model comprises nodes and edges in a directed acyclic graph (DAG). Nodes represent random variables, and edges represent their relationships. After identifying the nodes and connections, probabilities are assigned to nodes based on their interconnections and available information.

The chain rule, a core idea in BBNs, calculates the probability of complicated events from the probabilities of individual events, enabling jobs that involve reasoning with ambiguous or insufficient data. The chain rule mathematically states:

$$P(A, B, C...)=P(A|B)*P(B|C)*P(C|D)*...$$

Developing a BBN is a difficult process that requires understanding the relationships and dependencies between variables within a domain. There are two main ways to build a BBN: by using subject matter experts (SMEs) or data-driven methods. The SMEs approach depends on the domain knowledge and skills of SMEs to decide the structure of the BBN. When variable relationships are well understood and can be extracted from specialists, this method is useful. Unfortunately, it may not capture all underlying data connections and can be time-consuming and expensive. Data-driven methods, on the other hand, employ statistical techniques to understand the network topology from the data. When variable relationships are unclear or data is too complicated for experts, this method is preferable. One data-driven strategy, the Naive Bayes (NB) model, implies that all attributes are independent given the class. Yet, characteristics are rarely independent, and despite this, BBN models can still perform well by incorporating stronger conditional dependencies. Rather than creating a full BBN, augmented NB models integrate interdependence information between characteristics to improve classification performance. The Tree-Augmented Naive Bayes (TAN) model, which restricts the interaction between random variables, is one example. TAN preserves the fundamental structure of the NB model and incorporates all attributes when computing $P(C|A_1 A_2 ... A_N)$. Except for the class, each random variable is connected to another through a direct edge. This strategy minimizes computing complexity without compromising reliability or accuracy. Using a tree structure, TAN efficiently captures conditional independence in huge datasets. This algorithm begins by initializing an empty DAG (TAN) and then adding nodes for the target variable and all other variables in the dataset. The algorithm then adds edges from the target variable to all other nodes, calculates the mutual information (MI) for all variable pairs, and constructs a maximum spanning tree (MST) with the mutual information as edge weights. Lastly, it adds directed edges based on the proximity of the target variable to the MST nodes and returns the TAN model. In the TAN structure, the class variable has no parents, and each predictor is connected as a parent to the class variable along with a maximum of one other attribute. Consequently, the arc between two variables in the tree represents their probabilistic relationship (Fig. 4). It has been demonstrated that the TAN method outperforms other constraint-based structural learning algorithms, including Naive Bayes and Markov blanket (MB). The work of Korb and Nicholson (Korb & Nicholson, 2010) contains more information on Naive Bayes, TAN, and MB comparison. We chose the TAN model within the BBN framework for our research on human trafficking due to its superior performance compared to alternative methodologies. The primary objective of this model was to identify conditional dependencies among human trafficking predictors. Formal representation of parents for a variable x_i can be expressed as:

$$Pa_{x_i} = \left\{ C, x_{\delta(i)} \right\}$$

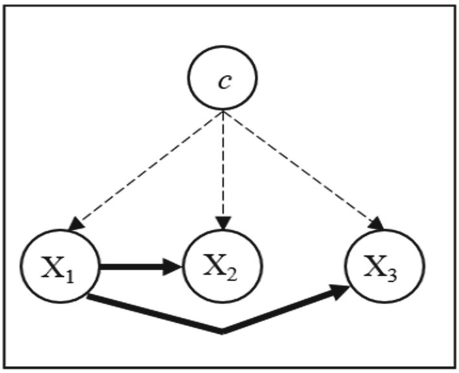


Fig. 4 A simple structure of the TAN

Here, the following notations are used:

Pa_{x_i}: The set of parents for variable x_i.
C: The class variable, which has no parents ($Pa_C = \varnothing$)
$x_{\delta(i)}$ Another variable in the set of parents for x_i

In this formulation, each predictor variable x_i has two parents—the class variable (C) and another variable $x_{\delta(i)}$, which is determined by the function δ(i).

3.5.1 Decision support tool–Inference simulator

Bayesian belief networks (BBNs) are dynamic, adaptive, and simple to update with new evidence, which makes them ideal for modeling evolutionary systems. For instance, a BBN can predict the association between weather, humidity, and an individual's chance of acquiring a cold. BBNs can provide a thorough picture of how variables influence one another by portraying variables as nodes and their relationships as edges in the network. Assigning probabilities to nodes enables the network to reveal the likelihood of particular events under given conditions. BBNs are crucial for comprehending conditional dependencies and relationships within a specific domain due to their capacity to be explained and their visual portrayal. They can be used to examine complicated systems with various variables in order to reveal their underlying relationships. The Inference Simulator is an instructional tool that allows user involvement with Bayesian network (BN) models and their comprehension. This decision support tool can recover and convey knowledge by visualizing, simulating, and analyzing information. It may also distribute interactive models via the web, making them accessible to the public. Its key benefit is its user-friendliness and intuitiveness, making it appropriate for users with varied degrees of technical skill. This enhances understanding of the model's dynamics and increases confidence in its predictions and inferences. The simulator essentially bridges the gap between human and artificial intelligence by providing a user-friendly and explicable interface for interacting with and comprehending Bayesian network models. It is a potent tool for knowledge discovery and decision-making, allowing users to easily explore and test a variety of situations and study the model's dynamics.

4 Results and discussion

4.1 Explainability with conventional models

This section discusses the explainability of more conventional methods, such as logistic regression and association rule analysis. While these methodologies provide valuable insights into pairwise relationships between variables, their capacity to reveal complex interactions between multiple factors is limited. Logistic regression is a well-established method for modeling the relationship between a binary outcome and a set of predictor variables; however, it may struggle to capture complex relationships when more than two variables are involved. Similarly, association rule analysis excels at identifying frequent patterns and robust rules in large datasets, but its emphasis on pairwise relations can obfuscate the big picture when multiple variables are involved. Therefore, even though these methods can provide some level of comprehension, they may fall short when it comes to elucidating the combined influence of multiple factors on a particular outcome.

4.1.1 Logistic regression results

All our variables are dichotomous (Table 2). The entire dataset with 11,368 records and 24 columns was admitted into the logistic regression model. With OLS, the typical goodness-of-fit measure is R^2 statistic corresponding to the proportion of variance explained by the predictors. With LR, however, several metrics could be used to assess the goodness-of-fit, including McFadden's *pseudo*-R^2 (ρ^2) which is computed using the log-likelihood of the null model (L_0) and that of the logistic regression (L_R) as $\rho^2 = 1 - (\frac{\ln(L_R)}{\ln(L_0)})$. Similar to R^2, with *pseudo*-R^2 statistics, values closer to zero correspond to less predictive power (McFadden, 1974). The LR model yielded a *pseudo*-R^2 of 0.83. We also compute the goodness-of-fit using two other popular *pseudo*-R^2 metrics (i) Cox and Snell (Cox & Snell, 1989) based on the log-likelihood of the model as 0.67, and (ii) Nagelkerke a modified version of Cox and Snell covering the full range from 0 to 1 as 0.91 (Nagelkerke, 1991). The overall p-value for the model with residual degrees of freedom ($rdf = 11,367$) and degrees of freedom ($df = 23$) is found to be near zero, $p < 0.001$,

For individual columns, Wald statistics indicate the statistical significance. It is the ratio of the square of the regression coefficient to the square of the SE. Table 4 presents coefficients, standard errors, and Wald statistics along with confidence intervals (CI) for regression coefficients (β). Our separation plot (Greenhill et al., 2011) (Fig. 5) also demonstrates a good "measure for refinement" for the LR model. We also perform receiver operator characteristic (ROC) to validate regression model results using the area under the curve (AUC). Both models' fit accuracy and ROC values are high (Accuracy = .959, Balanced Accuracy = 0.952, Sensitivity = 0.92, Specificity = 0.98, and AUC = 988). Collinearity could be an issue when there are a large number of predictors. We, therefore, compute variance inflation factors among the predictor variables and verify that all are less than the critical value of five (Fox & Monette, 1992).

The dependent variable in the LR is "Abuse Type" (FS = 1). For a similar prediction task, Cockbain and Bowers (2019) report a balanced accuracy rate of 76% for their classification and note that "Gender" and "Citizenship" as the most pronounced in terms of their predictive power for "Abuse type". Since Mexico and Ukraine citizenships were overrepresented and overly particular in our effort to draw broader generalizations, we purposely omitted the "Citizenship" variable. In parallel with their study, our analysis also confirms the gender disparity across trafficking types. However, our regression model is more inclusive and statistically performs significantly better. Our results are also consistent with previous research in that women are more prone to FS. According to the results, two key means of control, "psychoactive substances" and "excessive working hours," are the most exercised for FS and FL, respectively. Additionally, except for the "friendship" relationship type, the results indicate that all recruiter relationship types are significant. Closer relationships, such as those with an intimate spouse or family, are found to be more common among FS-cases. According to Table 4, most HT-victim minors are subject to FS (V2). Sexual abuse, psychoactive substances use, and family relationships are more prevalent among *FS* cases (variables: V9, V11, and V22).

4.1.2 Summary of the association rules analysis results

The logistic regression (LR) model is generated to distinguish between forced labor (FL) and forced sex (FS). However, it may not provide specific details about the pathways that contribute to each type of abuse. So, we employ association rule analysis (ARA) on two distinct datasets representing FL and FS cases to investigate these paths. Using the Apriori

Table 4 Logistic regression results

		β	SE	ZVals	Wald	CI (2.5%)	CI (97.5%)
Var#	Intercept	3.28***	0.25	12.87		2.78	3.78
V1	Gender (M = 1)	−3.46***	0.14	−24.69	609.5	−3.73	−3.18
V2	Minority status change	1.22***	0.35	3.51	12.28	0.54	1.90
V3	Debt bondage	−0.26†	0.15	−1.81	3.26	−0.55	0.02
V4	Takesearnings	−0.98***	0.12	−8.10	65.5	−1.22	−0.74
V5	Restricts financial access	−0.71**	0.27	−2.68	7.16	−1.24	−0.19
V6	Threats	−0.25	0.11	−0.28	0.08	−0.25	0.19
V7	Psychological abuse	−0.18	0.12	0.43	0.18	−0.19	0.29
V8	Physical abuse	−0.48***	0.12	−3.84	14.73	−0.72	−0.23
V9	Sexual abuse	0.66***	0.16	4.16	17.32	0.35	0.96
V10	False promises	−0.54***	0.12	−4.52	20.45	−0.77	−0.31
V11	Psychoactive substances	3.4***	0.29	11.74	137.7	2.83	3.97
V12	Restricts movement	−1.14	0.13	−0.73	0.53	−0.34	0.16
V13	Restricts medical care	−0.6*	0.27	−2.20	4.84	−1.14	−0.07
V14	Excessive working hours	−4.33***	0.28	−15.33	234.9	−4.89	−3.78
V15	Uses children	−0.69†	0.39	−1.78	3.17	−1.45	0.07
V16	Threat of law enforcement	−1.62***	0.18	−8.96	80.21	−1.98	−1.27
V17	Withholds necessities	−0.96***	0.14	−6.74	45.43	−1.23	−0.68
V18	Withholds documents	−1***	0.16	−6.05	36.59	−1.32	−0.67
V19	Other	0.5***	0.14	3.65	13.35	0.23	0.77
V20	Intimate partner	1.81***	0.30	5.99	35.85	1.21	2.40
V21	Friend	−0.86	0.25	−1.49	2.21	−0.86	0.12
V22	Family	1.09***	0.28	3.89	15.16	0.54	1.63
V23	Other	−1.5***	0.24	−6.25	39.06	−1.97	−1.03

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$; $^{\dagger}p < 0.1$

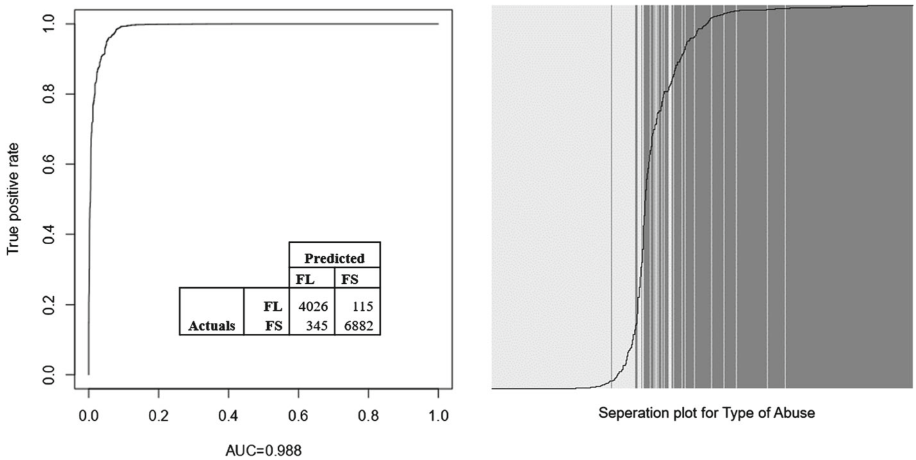


Fig. 5 ROC curve, confusion matrix, and separation plot for the logistic model

algorithm, the ARA method generates frequent item sets and extracts robust rules. By establishing confidence and support thresholds, we determine the top ten rules for each category of abuse, as shown in Appendix B. Appendix B demonstrates that FL paths are more distinct than FS paths and that psychological maltreatment is the only common method of control. Psychological abuse by an intimate partner is a significant factor in FS, whereas excessive working hours combined with fraudulent promises or non-familial relationships are prevalent patterns in FL. The most prevalent antecedents for FS and FL, respectively, are an intimate partner and excessive working hours. Incorporating these findings into human trafficking awareness campaigns could aid in alerting potential victims during recruitment, aid in self-identification, and ultimately aid in the fight against human trafficking. Understanding the distinct pathways leading to each type of abuse can aid in the development of more targeted and effective interventions, although the efficacy of such campaigns cannot be measured. Poelmans et al. (2012) also used concept hierarchy, a predecessor of association rules analysis. While their method is similar to ARA, their dataset, and therefore their resulting diagrams differ. The lattice diagrams produced in their study are similar to Fig. 10.

4.2 Explainability with BBN

BBNs have a distinct advantage over conventional machine learning models, such as logistic regression and association rule mining, due to their capacity to manage inference with multiple pieces of uncertain or contradictory evidence. This inherent ability to facilitate computations under uncertainty makes BNs ideally suited for a vast array of real-world applications, especially when investigating complex relationships between variables. Given the inadequacies of conventional methods in capturing complex connections and their emphasis on pairwise relationships, BBNs emerge as a superior instrument for providing a more comprehensive understanding. Our primary objective in this section is to develop an explanatory model rather than a strictly predictive one. While acknowledging the predictive efficacy of BBNs, we are primarily concerned with data interpretation using the inference model. By utilizing BBNs as our analytical framework, we can effectively circumvent the limitations of logistic regression and association rule mining. This enables us to delve deeper into the data and extract valuable insights, ultimately resulting in a more comprehensive analysis of the factors contributing to the various outcomes.

The performance metrics for the classification model for Forced Labor and Forced Sex demonstrate strong predictive capabilities. Both Forced Labor and Forced Sex have an accuracy of approximately 0.947, indicating that the model can accurately predict the vast majority of cases. The sensitivity for Forced Labor is 0.926, whereas the sensitivity for Forced Sex is 0.966, demonstrating that the model has a higher true positive rate for identifying Forced Sex cases. Similarly, Forced Labor has a specificity of 0.966 and Forced Sex has a specificity of 0.926, indicating that the model has a higher true negative rate for identifying non-Forced Labor cases. In addition, the overall AUC of 0.985 indicates that the model has an outstanding capacity for discrimination, which strengthens its ability to predict the type of abuse. More details are provided in Appendix C. In conclusion, these performance metrics indicate a robust and dependable model for separating Forced Labor and Forced Sex cases. While recognizing the predictive power of BBNs, we are primarily concerned with the explanation of variables using the inference model.

In our study of human trafficking, we utilized a tenfold stratified cross-validation technique, which required constructing ten distinct probabilistic inference models, each of which employed a distinct tenth of the entire dataset as the test set. Then, one of the ten models was chosen as the representative (model 3) for the construction of the illustrative model due to its performance closely resembling the average of all ten. This study uses entropy and mutual information to tackle uncertainty, as was described in the methodology section. These metrics quantify the uncertainty that exists in the probability distribution and identify the variable with the greatest predictive significance. In the context of our study on human trafficking, the degree of uncertainty is related to the predicted type of abuse. Figure 6 depicts a probabilistic graphical network in which values on nodes represent prior probabilities. Several factors should be examined in the context of the human trafficking BBN model, including

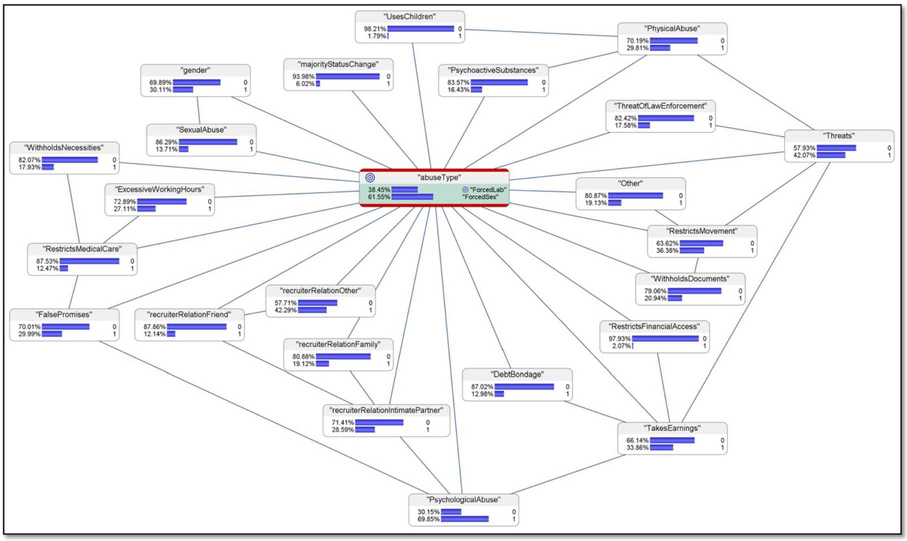


Fig. 6 Probabilistic graphical model for human trafficking factors: values on nodes represent prior probabilities, arcs show the dependencies between factors

arrows, interactions, and indirect effects. For example, "Takes Earnings" is associated with some factors, including Restrict Financial Access, Threats, Psychological Abuse, and Debt Bondage. Furthermore, there exists a direct relationship between Psychological Abuse and Intimate Partner Relations in the context of Recruiter Relations, as well as the phenomenon of False Promises. To provide an illustration utilizing the aforementioned information, let us contemplate a hypothetical situation wherein an individual's income is confiscated by their exploiter. The individual in question is subjected to multiple forms of abuse within the context of human trafficking, including limited financial resources, intimidation, psychological abuse, and being held in debt bondage. Additionally, the psychological abuse experienced by victims of trafficking may be linked to their association with the trafficker, who may have a personal relationship with them, as well as deceptive inducements that were employed to ensnare them in the trafficking scenario. The present illustration showcases the utility of the BBN model in discerning correlations and interdependencies between diverse variables and factors in a scenario of human trafficking. Through comprehending these interrelationships, it becomes feasible to formulate tactics for preempting and mitigating such instances.

According to sources, a BBN can function as an inference engine, allowing for thorough modeling of a given domain. The BBN model enables understanding of all associations within a given domain, and notably, facilitates accurate reasoning about a problem domain despite numerous uncertainties, while also enabling omnidirectional inference. The computation involved in deriving inference through simulation within a BBN is non-trivial, as noted in references (Nadkarni & Shenoy, 2001; Delen et al., 2020; Topuz et al., 2018; Hosseini & Ivanov, 2022; D'Urso et al., 2022). Algorithms have been implemented to perform the necessary tasks in the background, and are executed appropriately within the decision support tool, inference simulator. The simulator is a pedagogical tool that effectively retrieves and transmits knowledge within the Bayesian Network through the use of visualization, simulation, and analysis. As such, it serves as a bridge between human and artificial intelligence.

The Inference Simulator is utilized to disseminate interactive models through the internet, thereby enabling the public sharing of any Bayesian Network model constructed with a wider audience. The interface of the DSS tool is depicted in Fig. 7. Upon publication of a model through the simulator, users can experience various scenarios and their updated probabilities. The publicly available DSS simulator can be found here:

https://simulator.bayesialab.com/#!simulator/48726219478

To comprehend the interpretable conditional dependencies in the context of human trafficking, we created an example with omnidirectional inference and ran multiple simulations

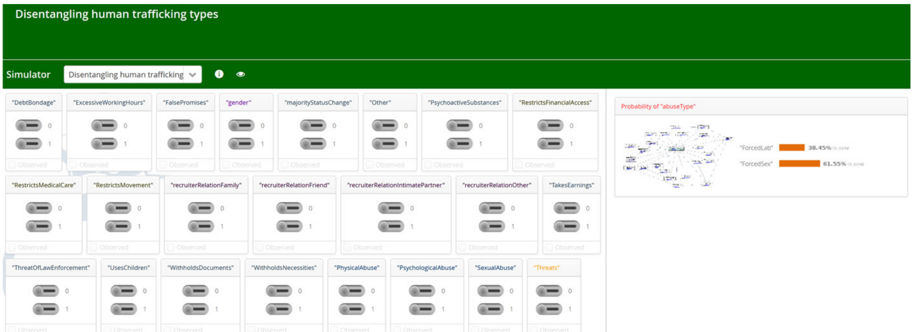
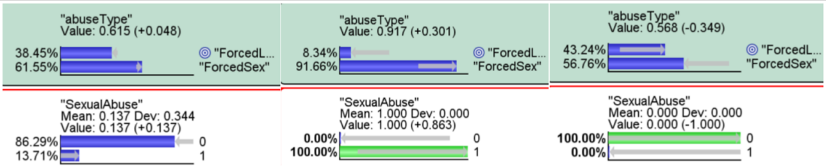
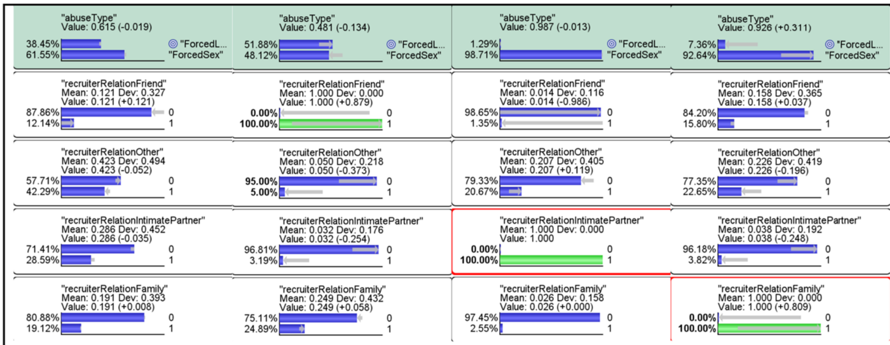


Fig. 7 Inference Simulator DSS tool screenshot

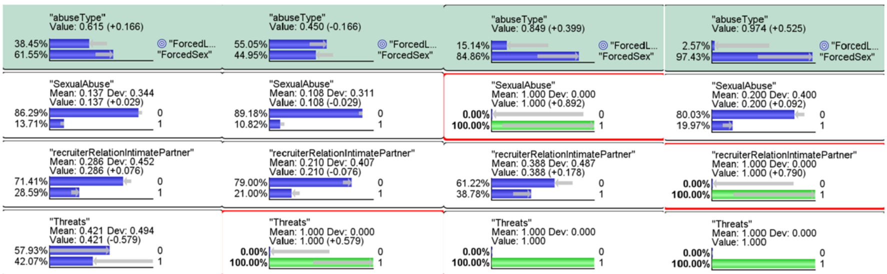
of the network's various components. Observing certain variables, for instance, can significantly alter the probability of distinct types of abuse. In one scenario (Fig. 8a.), the probability of the "Abuse type" being "Forced Sex" increases dramatically from 61.5 to 91.7% if "sexual abuse" is reported. When the recruiter is a friend (Fig. 8b), the probability that the abuse type is "Forced Labor" increases from 38.5 to 51.0%. However, if the recruiter is a romantic partner, the probability of "Forced Labor" decreases to 1.3%, while the probability of "Forced



a. Conditional probability update example1 (probability updates when "SexualAbuse" observed)



b. Conditional probability update example2 (probability updates when "recruiterRelation" observed)



c. Conditional probability update example3 (probability updates when "SexualAbuse", "Threats", and "recruiterRelation" observed)

Fig. 8 Conditional Probability Update: "Value" signifies the updated probability of "ForcedSex". The number in parentheses, located next to "Value", indicates the magnitude of increase in the "ForcedSex" probability following the observation. This highlights the belief update process in Bayesian belief networks

Sex" substantially increases to 98.7%. Similarly, if the recruiter is a family member, the probability of "Forced Sex" abuse increases to 92.6%. Another example (Fig. 8c) is if “threats’ is reported, the probability of the "Abuse type" being "Forced Labor" increases dramatically from 38.5 to 55.1%, however, if "sexual abuse" is reported or the recruiter is a romantic partner, the probability of the "Abuse type" being "Forced Labor" decreases dramatically to 15.1, and 2.6% respectively. Observers can observe these omnidirectional inference updates by modifying the values of specific variables and observing how the probabilities change accordingly. In our example, modifying a specific reported value, such as selecting a particular recruiter relationship, causes the distributions of the other variables to be recalculated, illustrating the interconnected nature of the factors underlying human trafficking.

4.3 Discussion of the results

Most quantitative work on human trafficking that relies on data has been restricted in terms of data coverage, volume, scope, and quality. Tyldum and Brunovskis (2005) use a Venn diagram to clearly demonstrate the coverage bias: trafficking victims lie in the intersection of migration and exploitation. Only a small subset of the trafficking victims are recognized by non-governmental organizations (NGOs) or social services. Moreover, trafficking cases make up just a tiny portion of this intersection, limiting data coverage. In such cases, the statistical remedies for non-response or sampling biases cannot be operationalized. The lack of adequate data volume reduces statistical power, resulting in poor representativeness, and diminishes robustness. This leads to an overreliance on anecdotal evidence, or sample size bias in analysis. Two studies: Cockbain and Bowers (2019), and Poelmans et al. (2012) sport two of the larger datasets. ($n = 2630$ cases and $n = 266{,}157$ reports yielding 4895 persons of interest, respectively). Needless to say, these numbers constitute a very small fraction of the trafficked population. Other studies such as Kiss et al. (2021), employed Bayesian Belief Networks (BBNs) in their study, albeit on a smaller sample of only 519 Nepalese refugees. Another limitation is related to the scope. Predictive models are intended to forecast a single variable of interest. As a result, the model’s performance is solely assessed in relation to that one variable of interest. Variables related to demographics, recruitment and exploitation methods, trafficking routes or networks, risk factors and vulnerabilities, victim characteristics, demand-side factors can be incorporated into the prediction models as dependent or independent variable depending on the scope selection. Often, the availability of the variables dictates or steers the study in a certain direction, as in this and the three other studies listed above. This results in availability bias. Finally, there is the issue of data quality. The data might be fragmented across multiple sources such as law enforcement and NGOs. Furthermore, due to its clandestine nature and reluctance of the victims to come forward data reliability and validity are concerns. Even with the k-anonymization, some victims do not wish their data disclosed implying additional sampling bias.

This study addresses and relieves some of these data-related restrictions. First, in terms of data volume, it relies on a larger sample size. Second, in terms of data coverage, our dataset offers a wider selection of variables (than Cockbain and Bowers (2019)). Hence employs a model that demonstrates the predictive efficacy of variables other than gender. Third, in terms of scope, omnidirectional inferencing in Bayesian belief networks updates probabilities by modifying the values of specific, case-based variables accordingly, even in the absence of a specific target variable. Lastly, in terms of data quality, the study relies on arguably the most reliable publicly available data source. Additionally, we believe that making the web simulator publicly available also enables the stakeholders to perform their own analysis.

This aspect further distinguishes this study from previous ones. We believe that future studies should prioritize the creation of such decision-making tools.

5 Conclusion

We believe that this study demonstrates a significant application of analytics in combating HT. We confirm and challenge some of the findings in the extant literature and call for better prevention strategies and more research on HT. The study contributes to the literature on several fronts. First, our model goes beyond validating some preexisting statements about HT, such as the predictive power of the “Gender” variable (Cockbain & Bowers, 2019). Using BBNs our study reveals many additional significant predictors such as controlling using “Excessive working hours” and “Psychoactive substances”. Furthermore, when variable relationships are unclear or data is too complicated for experts, BNNs provide an adaptive framework that could be updated with new evidence. Through the creation of a BBN decision tool, subject matter experts may update the network with fresh evidence and acquire additional insight depending on the task. Democritus stressed the significance of comprehending causation in complicated systems. In the difficult topic of human trafficking, BBN can be utilized to model probability and show intricate linkages. The suggested methodology produces understandable structures and leverages theory-based, data-driven BBN models as "inference engines" for automated inference and prediction. This approach can help us better understand the elements that contribute to human trafficking and inspire targeted solutions to tackle this multidimensional problem. One possible way of expanding our research is by annexing resource allocation into these identified pathways. For instance, given the evidence and/or prior probabilities of identified potential pathways leading to HT, how successfully might these pathways be utilized in awareness campaigns? Or, how should the available resources be allocated given the respective cost figures, as well as confidence and support for these paths?

This study also has several shortcomings, most of which are attributable to domain-related issues, particularly data availability. According to the literature, there could be around 25 million HT cases at any given moment. Therefore, the dataset is remarkably constrained, including just a tiny portion of the unknown size of all HT cases. One data-related concern is reporting bias. Most of the cases in the dataset were gathered in the U.S., either via hotline or through face-to-face case management. Another data-related issue is coverage bias. For instance, in terms of citizenship of HT victims, the Philippines, Mexico, and Ukraine are over-represented (62% of the records), and more than three-quarters of exploits took place in the U.S. Furthermore, sampling bias (i.e. there are more FS instances in our dataset, but reliable sources estimate a greater prevalence for FL) also exist. In addition, the de-identified dataset still omits the records if the victims don’t give informed consent for using their data the rights and dignity of victims. However, these are not limitations specific to our research. For instance, a recent study using data from the UK (Cockbain & Bowers, 2019) reported that the victims from nine nationalities (Nigerian, Romanian, British, Albanian, Polish, Vietnamese, Slovakian, Hungarian, Czech, Lithuanian) accounted for nearly 3 quarters of the overall sample. They also drew attention to the risk of bias. Another data-related issue is missing values. The original dataset before filtering included 95,739 rows but with 769 complete cases only. 57% of the data were missing. While we did not perform any missing value imputation, we had to make certain assumptions, such as interpreting missing values in variables linked to “means of control” or “relationship” as indicators of their absence (zero rather than NA).

Another issue is that the dataset's temporal resolution prevents research on the temporal element of HT. The HT cases typically require long processing times (a median of 61 days, according to Cockbain and Bowers (2019)) over which some of the data about the cases might change. Last but not least, a considerable number of HT cases associated with state-imposed FL, such as in Xinjiang (Paton & Ramzy, 2020), cannot yet be examined using analytical techniques. These limitations (or biases) necessitate assessing the CTDC dataset, to understand its drawbacks and to provide a critical assessment of research studies that rely on it.

Funding No funding was received to assist with the preparation of this manuscript.

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Appendix A

The summary definitions of the variables according to the CTDC data codebook V7 (CTDC, 2021) and other additional definitions used.

Variable	Explanation
Human trafficking	The recruitment, transportation, transfer, harboring or receipt of persons by means of threat, use of force, or other forms of coercion, abduction, fraud, deception, abuse of power or vulnerability for the purpose of exploitation. Exploitation includes sexual exploitation, forced labor or services, slavery or practices similar to slavery, servitude or the removal of organs. (UNODC, 2004)
Abse type: Forced sex	The victim was trafficked with the purpose of exploiting them for sexual services, which may include prostitution
Abuse type: Forced labor	The victim was coerced or threatened to perform work or service against their will, without offering themselves voluntarily, with the exclusion of sexual services
Gender	The victim's gender expression or condition of being male or female or non-binary

	Variable	Explanation
	Minority status change	The dataset provides information on the individual's minority status when they entered the trafficking process, which is recorded under "minorityEnd". It also indicates whether the victim was under the age of 18 or over at the time they were registered and assisted, which is noted under "minorityStatus". In cases where the two statuses differ, it can be inferred that the exploitation occurred when the victim was a minor
Means of control	Threats	Whether the victim was subjected to explicit or implicit communication by their exploiter(s) intending to cause harm or loss to the victim. It does not encompass threats of reporting or contacting law enforcement
	Restricts movement	Whether the exploiter(s) subjected the victim to physical or social isolation, confinement, or restriction of movement
	Takes earnings	Whether the victim has experienced a situation where the exploiters have taken his/her remuneration to control him/her
	False promises	Whether the victim was deceived or misled by their exploiter(s) using false promises or misrepresentation, resulting in their entry into an exploitative situation
	Psychological abuse	Whether the victim was subjected to emotionally abusive, deceptive, or manipulative tactics by their exploiter(s) to influence them, which may involve actions such as name-calling, verbal abuse, public humiliation, manipulation of power dynamics, shaming, or blaming
	Physical abuse	Whether the victim was subjected to physical harm, injury, disability, trauma, or even death by their exploiter(s). Note that physical acts of sexual abuse are treated separately and categorized in the variable labeled "Sexual Abuse"
	Excessive working hours	Whether the victim was compelled to work a substantial number of hours beyond what was promised to them, which may involve overtime, late or unusual shifts, or overnight hours

Variable	Explanation
Withholds documents	Whether the victim's exploiter(s) restricted, controlled, or limited their access to important documents, including but not limited to their passport, immigration documents
Withholds necessities	Whether the victim was denied or threatened with the denial of basic living necessities by their exploiter(s), including but not limited to food, shelter, water, hygiene, appropriate clothing, or necessary items
Threat of law enforcement	Whether the victim was subjected to explicit or implicit communication by their exploiter(s) intending to contact or involve law enforcement or other relevant authorities, including immigration authorities, with the aim of causing harm or negative consequences to the victim
Psychoactive substances	Whether the exploiter(s) coerced the victim into substance abuse, supplied them with substances to render them compliant or to influence their behavior, or exploited an existing substance abuse issue
Sexual abuse	Whether the victim was subjected to any form of non-consensual sexual contact by their exploiter(s) as a means of controlling them, rather than for the purpose for which they were trafficked. It also includes coercive behavior that impairs the individual's reproductive rights. The term may be confounded with "sexual exploitation" (or FS in our case) where the victim is led to believe that the victim is led to believe the situation is not as abusive (UNODC, 2020)
Debt bondage	Whether the victim was coerced into working to repay a fabricated or perceived debt. This definition is outlined in the United Nations' 1956 Supplementary Convention on the Abolition of Slavery
Restricts medical care	Whether the exploiter(s) restricted the victim's access to medical or health services, which may involve withholding necessary medical care or treatment or controlling access to such treatment

	Variable	Explanation
	Restricts financial access	Whether the victim was prohibited or restricted from accessing their own personal finances or necessary daily living funds by their exploiter(s)
	Uses children	Whether the exploiter limited the victim's access to their children as a form of control
	Other	Whether the exploiter used a tactic intended to establish or sustain power and control over the victim, which cannot be reasonably classified under any of the other categories
Recruiter relation	Intimate partner	Whether a person who initially enticed or obtained the victim into the situation of exploitation was one with whom the individual has identified having a current or former romantic relationship
	Other	Whether a person who initially enticed or obtained the victim into the situation of exploitation was a person with whom the individual had any other notable relationship that cannot reasonably fit into previous categories. This may include, but is not limited to, labor brokers, contractors, formal employers, or smugglers
	Family	Whether a person who initially enticed the victim into the situation of exploitation was one with whom the individual was connected biologically, through marriage, or a person who the individual has identified as having been their current or former custodian or guardian
	Friend	Whether the person who recruited the victim into an exploitative situation was someone known to the individual, excluding romantic partners, family members, or other formal relationships

Appendix B: The summary of Association Analysis

The LR model is built to distinguish FL from FS. Regardless of the goodness of fit of LR model, it may offer little to no information about specific pathways leading to each type of abuse. At this point, we focus on investigating these paths. FS and FL are collectively exhaustive events for the "Abuse Type" variable. The variable "Abuse Type" encodes FS as one and FL as zero. Using our dataset in its present form, the extracted rules will only relate to FS cases. The ARA literature advises mining "negative association rules". This allows for the investigation of relationships between both present (one) and absent (zero) items. These negative association rules, on the other hand, consider the existence and absence in both the antecedent and consequent parts of the rule. In our analysis, we are only interested in evaluating the dichotomy relevant to the rule’s consequent component only. In lieu of mining the negative association rules, we generate two variations of our dataset using alternate "Abuse Type" representations: (i) FS variant: the consequent (Abuse Type) for FS records is set to 1; (ii) FL variant: the consequent (Abuse Type) for FL records is set to 1. These two datasets, one for each type of abuse, are then loaded into our ARA engine individually. All variables in our dataset are dichotomous and require no further data pre-processing. ARA is typically performed in two steps: generation of frequent itemsets, and extraction of strong rules. There are a variety of algorithmic options for producing frequent item sets. While these algorithms vary in their performance (Zheng et al., 2001) in terms of execution time, needed database scans and completeness, our dataset is regarded to be small. Therefore, we select one of the most popular algorithms, Apriori (Agrawal & Srikant, 1994) with Christian Borgelt’s implementation. Ghafari and Tjortjis (2019) provide interested readers a comprehensive review of various ARA techniques.

In our analysis, we perform ARA once for each of the two formed datasets. Furthermore, we limit the rule extraction to the consequent variable (“Abuse Type”), as we are primarily interested in the paths (in the form of rules) that lead to each abuse type. We, therefore, fix the consequent in ARA analysis as FS and FL for each dataset. By trial-and-error, we specify confidence and support thresholds of 0.06 and 0.8 for each dataset and report top 10 rules for inspection (Table 5, Figs. 9 and 10).

Figure 10 suggests that paths leading to FL are more distinct (thicker lines indicating slightly stronger support) than those leading to FS. "Psychological Abuse" is the sole means of control shared by both types. The intimate partner’s "psychological abuse" seems to be one of the key aggravators of FS. “Excessive working hours” coupled with either “false promises” or relationships other than family, intimate partner and friends seem to be the most prevalent patterns leading to FL. The bulk of the arrows (Fig. 10) emanate from “intimate partner” and “excessive working hours” make these the two most common antecedents of FS and FL. While the anti-human trafficking campaigns’ effectiveness have not been measured (Szablewska & Kubacki, 2018), and most work needs to be done in that department, HT-awareness campaigns are arguably the most effective tools to detect and deter HT (Konrad et al., 2017). In accordance with O’Brien (2016), we recognize that the narrow construction of HT restricts our capacity to combat it. However, we believe that, within the scope of scholarly work, our framework might be implemented by integrating these paths into campaign designs. Our findings could be incorporated in these campaigns for a variety of reasons including, but not limited to warning victims during recruiting or the early stages of HT, as well as helping them in self-identification.

Table 5 Association rules results for Forced Sex (a) and Forced Labor (b)

Rule #	Antecedent	Support	Confidence	Lift	Count
(a) Association rules for forced sex					
1	{Psychoactive Substances, Intimate Partner}	0.08	1.00	1.62	882
2	{Psychological Abuse, Intimate Partner}	0.25	0.99	1.60	2887
3	{Other, Intimate Partner}	0.06	0.99	1.60	723
4	{Physical Abuse, Intimate Partner}	0.12	0.99	1.60	1325
5	{Control Threats, Intimate Partner}	0.09	0.98	1.59	1057
6	{Restricts Movement, Intimate Partner}	0.08	0.98	1.59	875
7	{Psychological Abuse, Psychoactive Substances}	0.10	0.96	1.57	1179
8	{Psychological Abuse, Family}	0.16	0.95	1.54	1808
9	{Physical Abuse, Psychoactive Substances}	0.06	0.95	1.54	705
10	{Control Threats, Psychoactive Substances}	0.06	0.95	1.54	712
(b) Association rules for forced labor					
11	{Gender, Excessive Working Hours}	0.20	1.00	2.60	2328
12	{Gender, Takes Earnings}	0.20	1.00	2.59	2235
13	{False Promises, Excessive Working Hours}	0.19	1.00	2.59	2205
14	{Excessive Working Hours, Other Relationship}	0.22	1.00	2.59	2527
15	{Takes Earnings, Excessive Working Hours}	0.21	1.00	2.59	2412
16	{Psychological Abuse, Excessive Working Hours}	0.21	0.99	2.59	2331
17	{Gender, Other Relationship}	0.22	0.99	2.57	2514
18	{Takes Earnings, False Promises}	0.20	0.95	2.48	2244
19	{Gender, Psychological Abuse}	0.19	0.92	2.39	2204
20	{Takes Earnings, Other Relationship}	0.21	0.90	2.35	2389

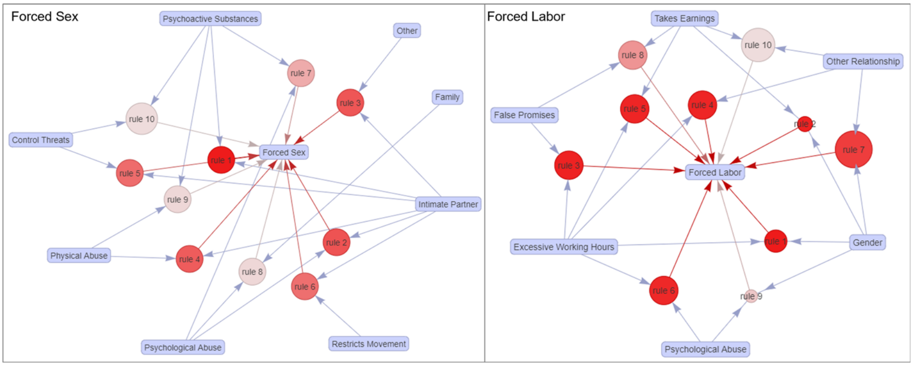


Fig. 9 Top 10 rules according to “lift” measure for FS and FL

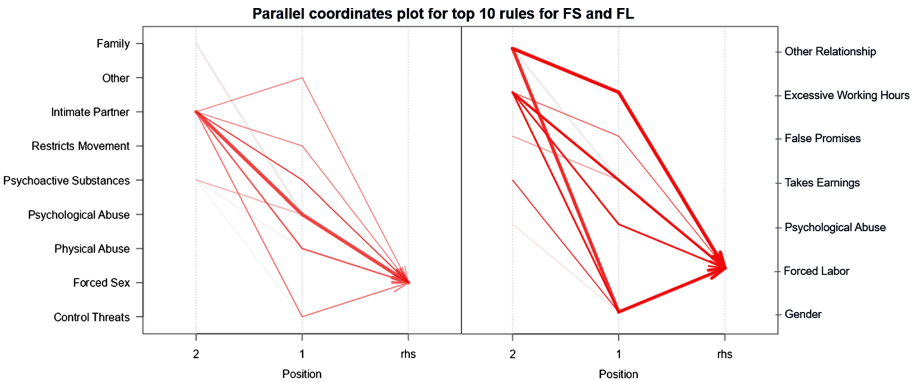


Fig. 10 Parallel coordinates plot for top 10 rules according to “lift” measure for FS (left) and FL (right)

Appendix C

Table 6 presents the tenfold cross-validated results for four measures, including their mean, minimum, maximum, and best network number. The mean precision is 94.60%, the range is 93.48–95.28%, and the finest network is #5. The top network has a mean reliability of 94.94%, ranging from 94.07 to 95.53%, with the best network being #9. The ROC index has a mean value of 98.51%, with values ranging from 97.61 to 99.05%; the top network is

Table 6 Cross-validated BBN results

Measure	Mean	Min	Max	Best network#
Precision	94.60%	93.48%	95.28%	5
Reliability	94.94%	94.07%	95.53%	9
ROC index	98.51%	97.61%	99.05%	7
Binary log-loss	0.2862	0.2325	0.4016	3

#7. Lastly, binary log-loss has a mean value of 0.2862, a minimum value of 0.2325, and a maximum value of 0.4016, and the best network for this metric is #3.

References

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of 20th international conference on very large data bases, VLDB* (Vol. 1215), pp. 487–499.

Alvari, H., Shakarian, P., & Snyder, J. E. (2017). Semi-supervised learning for detecting human trafficking. *Security Informatics, 6*(1), 1–14.

Alvari, H., Shakarian, P., & Snyder, J. K. (2016). A non-parametric learning approach to identify online human trafficking. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*, pp. 133–138.

Andrews, S., Brewster, B., & Day, T. (2018). Organised crime and social media: A system for detecting, corroborating and visualising weak signals of organised crime online. *Security Informatics, 7*(1), 1–21.

Bales, K. (2007). What predicts human trafficking? *International Journal of Comparative and Applied Criminal Justice, 31*(2), 269–279.

Burbano, D., & Hernández-Alvarez, M. (2018a). Illicit, hidden advertisements on Twitter. In *2018aInternational Conference on eDemocracy & eGovernment (ICEDEG)* (pp. 317–321). IEEE.

Burbano, D., & Hernández-Alvarez, M. (2018b). Illicit, hidden advertisements on Twitter. In 2018b International Conference on eDemocracy & eGovernment (ICEDEG) (pp. 317–321). IEEE.

Caltagirone, S. (2017). The Human Trafficking Kill Chain, Retrieved from: https://www.globalemancipation.ngo/wp-content/uploads/2021/11/Human-Trafficking-Kill-Chain-2.pdf, Accessed: April 4, 2023.

Chen, Q., De Arteaga, M., & Herlands, W. (2015). Canonical autocorrelation analysis and graphical modeling for human trafficking characterization. Carnegie Mellon University.

Cockbain, E., & Bowers, K. (2019). Human trafficking for sex, labour and domestic servitude: How do key trafficking types compare and what are their predictors? *Crime, Law and Social Change, 72*, 9–34.

Cockbain, E., Bowers, K., & Dimitrova, G. (2018). Human trafficking for labour exploitation: The results of a two-phase systematic review mapping the European evidence base and synthesising key scientific research evidence. *Journal of Experimental Criminology, 14*, 319–360.

Cockbain, E., Brayley, H., & Laycock, G. (2011). Exploring internal child sex trafficking networks using social network analysis. Policing: *A Journal of Policy and Practice*, 5(2), 144–157.

Couch, R. (2015), "The Average Hipster Employs 27 Slaves Each Day. Here's How To Change That", *Huffington Post,* https://www.huffpost.com/entry/hipsters-free-slaves_n_7172402?ncid=tweetlnkushpmg00000058. Retrieved: 08/08/2022

Cox, D. R., & Snell, E. J. (1989). *The Analysis of Binary Data* (2nd ed.). Chapman and Hall.

Crane, A., (2013) "Modern slavery as a management practice: Exploring the conditions and capabilities for human exploitation". *Academy of Management Review*, 38 (1), (pp. 49–69).

CTDC, (2021). "Counter Trafficking Data Collaborative (CTDC) K-anonymized Data Codebook", https://www.ctdatacollaborative.org/sites/g/files/tmzbdl2011/files/CTDCk_codebook_v7_1.pdf. Retrieved: 03/29/2023.

CTDC, (2022), "The Counter Trafficking Data Collaborative (CTDC): Global Data Hub on Human Trafficking" https://www.ctdatacollaborative.org/. Accessed: 08/08/2022.

da Silva Santos, M., Ladeira, M., Van Erven, G. C., & da Silva, G. L. (2019a). Machine learning models to identify the risk of modern slavery in Brazilian cities. In *2019a18th IEEE International Conference On Machine Learning And Applications (ICMLA)* (pp. 740–746). IEEE.

da Silva Santos, M., Ladeira, M., Van Erven, G. C., & da Silva, G. L. (2019b). Machine learning models to identify the risk of modern slavery in Brazilian cities. In 2019b 18th *IEEE International Conference On Machine Learning And Applications (ICMLA)* (pp. 740–746). IEEE.

Delen, D., Topuz, K., & Eryarsoy, E. (2020). Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition. *European Journal of Operational Research, 281*(3), 575–587.

Diaz, M., & Panangadan, A. (2020). Natural language-based integration of online review datasets for identification of sex trafficking businesses. In *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)* (pp. 259–264). IEEE.

Dimas, G. L., Khalkhali, M. E., Bender, A., Maass, K. L., Konrad, R., Blom, J. S., Zhu, J., & Trapp, A. C. (2020). Estimating Effectiveness of Identifying Human Trafficking via Data Envelopment Analysis. *arXiv preprint* arXiv:2012.07746.

Dimas, G. L., Konrad, R. A., Lee Maass, K., & Trapp, A. C. (2022). Operations research and analytics to combat human trafficking: A systematic review of academic literature. *PLoS ONE, 17*(8), e0273708.
D'Urso, P., De Giovanni, L., & Vitale, V. (2022). A Bayesian network to analyze basketball players' performances: A multivariate copula-based approach. Annals of Operations Research, 1–22.
Dubrawski, A., Miller, K., Barnes, M., Boecking, B., & Kennedy, E. (2015). Leveraging publicly available data to discern patterns of human-trafficking activity. *Journal of Human Trafficking, 1*(1), 65–85.
Farrell, A., Bright, K., de Vries, I., Pfeffer, R., & Dank, M. (2020). Policing labor trafficking in the United States. *Trends in Organized Crime, 23*, 36–56.
Feehs, K., A. Currier. (2020). 2019 Federal Human Trafficking Report. Technical report, *The Human Trafficking Institute*.
Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *JASA, 87*, 178–183.
Ganter, B., Stumme, G., & Wille, R. (Eds.). (2005). Formal concept analysis: foundations and applications (Vol. 3626). Springer, Cham.
Ghafari, S. M., & Tjortjis, C. (2019). A survey on association rules mining using heuristics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9*(4), e1307.
Giommoni, L., & Ikwu, R. (2021). Identifying human trafficking indicators in the UK online sex market. *Trends in Organized Crime*, pp.1–24.
Greenhill, B., Ward, M. D., & Sacks, A. (2011). The separation plot: A new visual method for evaluating the fit of binary models. *American Journal of Political Science, 55*(4), 991–1002.
Hachey, L. M., & Phillippi, J. C. (2017). Identification and management of human trafficking victims in the emergency department. *Advanced Emergency Nursing Journal, 39*(1), 31–51.
Hosseini, S., & Ivanov, D. (2022). A new resilience measure for supply networks with the ripple effect considerations: A Bayesian network approach. *Annals of Operations Research, 319*(1), 581–607.
Hoyle, C., Bosworth, M., & Dempsey, M. (2011). Labelling the victims of sex trafficking: Exploring the borderland between rhetoric and reality. *Social & Legal Studies, 20*(3), 313–329.
Human Trafficking Institute, (2021), "Federal Human Trafficking Report", Available at: https://traffickinginstitute.org/wp-content/uploads/2022/09/2021-Federal-Human-Trafficking-Report-WEB-1.pdf. Accessed: 09/04/2023.
ILO, (2014). " International Labour Organization: Profits and poverty: The economics of forced labour" (PDF). https://www.ilo.org/global/topics/forced-labour/publications/WCMS_243391/lang--en/index.htm . Accessed: 08/08/2022.
ILO, (2022), "Report: Global Estimates of Modern Slavery: Forced Labour and Forced Marriage", September 12, 2022. https://www.ilo.org/global/topics/forced-labour/publications/WCMS_854733/lang-en/index.htm. Accessed: March 30, 2023.
Kejriwal, M., & Szekely, P. (2017a). Information extraction in illicit web domains. In *Proceedings of the 26th international conference on world wide web*, pp. 997–1006.
Kejriwal, M., & Szekely, P. (2022). Knowledge graphs for social good: An entity-centric search engine for the human trafficking domain. *IEEE Transactions on Big Data, 8*(3), 592–606.
Keskin, B. B., Bott, G. J., & Freeman, N. K. (2021). Cracking sex trafficking: Data analysis, pattern recognition, and path prediction. *Production and Operations Management, 30*(4), 1110–1135.
Kiss, L., Fotheringhame, D., Mak, J., McAlpine, A., & Zimmerman, C. (2021). The use of Bayesian networks for realist evaluation of complex interventions: Evidence for prevention of human trafficking. *Journal of Computational Social Science, 4*(1), 25–48.
Konrad, R. A. (2019). Designing awareness campaigns to counter human trafficking: An analytic approach. *Socio-Economic Planning Sciences, 67*, 86–93.
Konrad, R. A., Trapp, A. C., Palmbach, T. M., & Blom, J. S. (2017). Overcoming human trafficking via operations research and analytics: Opportunities for methods, models, and applications. *European Journal of Operational Research, 259*(2), 733–745.
Korb, K. B., & Nicholson, A. E. (2010). *Bayesian Artificial Intelligence*. Chapman & Hall/CRC.
Libaque-Saenz, C. F., Lazo, J., Lopez-Yucra, K. G., & Bravo, E. R. (2017). Could Machine Learning Improve the Prediction of Child Labor in Peru?. In *Annual International Symposium on Information Management and Big Data*, pp. 15–30.
Maass, K. L., Trapp, A. C., & Konrad, R. (2020). Optimizing placement of residential shelters for human trafficking survivors. *Socio-Economic Planning Sciences, 70*, 100730.
McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105–142). Academic Press.
Nadkarni, S., & Shenoy, P. P. (2001). A Bayesian network approach to making inferences in causal maps. *European Journal of Operational Research, 128*(3), 479–498.
Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78*(3), 691–692.

National Crime Agency. (2018). *Modern slavery and human trafficking: National Referral Mechanism Statistics Annual Report 2017*. National Crime Agency.
O'Brien, E. (2016). Human trafficking heroes and villains: Representing the problem in anti-trafficking awareness campaigns. *Social & Legal Studies, 25*(2), 205–224.
Office to Monitor and Combat Trafficking in Persons. (2019) 2019 trafficking in persons report. Report, United States Department of State. URL https://www.state.gov/wp-content/uploads/2019/06/2019-Trafficking-in-Persons-Report.pdf. Accessed: August 15,2022.
Okech, D., Choi, Y. J., Elkins, J., & Burns, A. C. (2018). Seventeen years of human trafficking research in social work: A review of the literature. *Journal of Evidence-Informed Social Work, 15*(2), 103–122.
Paton, E., Ramzy, A. (2020), "Coalition Brings Pressure to End Forced Uighur Labor". *New York Times, 10 August 2020.* Accessed: 08/08/2022.
Poelmans, J., Elzinga, P., Ignatov, D. I., & Kuznetsov, S. O. (2012). Semi-automated knowledge discovery: Identifying and profiling human trafficking. *International Journal of General Systems, 41*(8), 774–804.
Ramchandani, P., Bastani, H., & Wyatt, E. (2021). Unmasking human trafficking risk in commercial sex supply chains with machine learning. Available at SSRN 3866259.
Schauer, E. J., & Wheaton, E. M. (2006). Sex trafficking into the United States: A literature review. *Criminal Justice Review, 31*(2), 146–169.
Sheinis, D. (2012). The links between human trafficking, organized crime, and terrorism. *American Intelligence Journal, 30*(1), 68–77.
Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10*(05), 557–570.
Sweileh, W. M. (2018). Research trends on human trafficking: A bibliometric analysis using Scopus database. *Globalization and Health, 14*, 1–12.
Szablewska, N., & Kubacki, K. (2018). Anti-human trafficking campaigns: A systematic literature review. *Social Marketing Quarterly, 24*(2), 104–122.
Szakonyi, A., Chellasamy, H., Vassilakos, A., & Dawson, M. (2021). Using technologies to uncover patterns in human trafficking. In ITNG 2021 18th International Conference on Information Technology-New Generations (pp. 497–502). Cham: Springer International Publishing.
The U.S. Department of Justice (2020). "Justice Department Awards Nearly $101 Million to Combat Human Trafficking", https://www.justice.gov/opa/pr/justice-department-awards-nearly-101-million-combat-human-trafficking#:~:text=Justice%20Department%20Awards%20Nearly%20%24101%20Million%20to%20Combat%20Human%20Trafficking,-The%20Department%20of . Accessed: 08/08/2022.
Topuz, K., Uner, H., Oztekin, A., & Yildirim, M. B. (2018). Predicting pediatric clinic no-shows: A decision analytic framework using elastic net and Bayesian belief network. *Annals of Operations Research, 263*, 479–499.
Tyldum, G., & Brunovskis, A. (2005). Describing the unobserved: Methodological challenges in empirical studies on human trafficking. *International Migration, 43*(1–2), 17–34.
U.S. Department of State, (2022a). "Trafficking in Persons Report", *p.62*, https://www.state.gov/wp-content/uploads/2022a/10/2022a1020-2022-TIP-Report.pdf. Accessed: 08/04/2023.
U.S. Department of State, (2022b). "Trafficking in Persons Report", *p.38*, https://www.state.gov/wp-content/uploads/2022b/10/2022b1020-2022-TIP-Report.pdf. Accessed: 08/04/2023.
UN. (2015). *United Nations: Sustainable development goals: 17 goals to transform our world*. United Nations.
United Nations (2018). "United Nations Convention against Transnational Organized Crime and the Protocols Thereto", URL: https://www.unodc.org/unodc/en/organized-crime/intro/UNTOC.html. Accessed: 08/08/2022.
UNODC (2004), "United Nations convention against transnational organized crime and the protocols thereto", https://www.unodc.org/documents/treaties/UNTOC/Publications/TOC%20Convention/TOCebook-e.pdf, Accessed: 3/29/2023.
UNODC (2020), "Global Report on Trafficking in Persons 2020", https://www.unodc.org/documents/data-and-analysis/tip/2021/GLOTiP_2020_15jan_web.pdf. Accessed: 08/08/2022.
UNSC, (2017), "United Nations Security Council S/RES/2388- Fight against human trafficking in Libya: EU sanctions six human traffickers and smugglers in line with UN decision". Retrieved From https://www.consilium.europa.eu/en/press/press-releases/2018/06/14/fight-against-humantrafficking-in-libya-eu-sanctions-six-humantraffickers-and-smugglers-in-line-with-undecision/
Wheaton, E. M., Schauer, E. J., & Galli, T. V. (2010). Economics of human trafficking. *International Migration, 48*(4), 114–141.
Whitman, A., & Gray, D. H. (2015). Transnational Human Trafficking. *Global Security Studies*, *6*(3).
Zheng, Z., Kohavi, R., & Mason, L. (2001). Real world performance of association rule algorithms. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 401–406.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.