ANALYZING HUMAN TRAFFICKING DATA

A Machine Learning Approach

September 2019

Capstone Course Report

IOM
UN MIGRATION

University of Zurich[UZH]

The opinions expressed in the report are those of the authors and do not necessarily reflect the views of the International Organization for Migration (IOM). The designations employed and the presentation of material throughout the report do not imply the expression of any opinion whatsoever on the part of IOM concerning the legal status of any country, territory, city or area, or of its authorities, or concerning its frontiers or boundaries.

The data analysis was conducted from February to June 2018. All data, figures and citations refer to information available from 2002 until 2017.

Details provided by identified and assisted victims have been changed to prevent identification.

ACKNOWLEDGEMENTS

This report was written by Janine Albinez, Patrik Aus der Au, Saro Gibilisco, Théoda Woeffray and Jenny Yin under the guidance and supervision of Dr. Judith Spirig (University of Zurich), Eliza Galos (IOM) and Prof. Marco Steenbergen (University of Zurich).

The report was produced in the context of a joint project between the International Organization for Migration (IOM) and the Department of Political Science at the University of Zurich (UZH). The project was carried out by Master students within a Capstone Course with the aim to use students' research and quantitative data analysis skills to support the evidence base for IOM's counter trafficking efforts. In return, the students had the opportunity to learn about migration and human trafficking and the identification of trafficking victims. This publication was made possible by accessing IOM's global database on victims of human trafficking.

The authors are very thankful to Harry Cook and Mathieu Luciano for their review. They also thank to Damiano Toffano for his presentation to the authors on assistance to victims of human trafficking, and to Anh Nguyen who facilitated the final presentation on this work in Geneva, at IOM headquarters.

ACRONYMS

CENA	Central Asia
DTM	Displacement Tracking Matrix
EEA	European Economic Area
EU	European Union
FMP	Flow Monitoring Point
FMS	Flow Monitoring Survey
ILO	International Labour Organization
IOM	International Organization for Migration
MENA	Middle East & North Africa
SEE	South-Eastern Europe
UNICEF	United Nations International Children's Emergency Fund
UNODC	United Nations Office on Drugs and Crime

INTRODUCTION

Human trafficking is a human rights violation, a pervasive and destructive phenomenon in our global society, often underestimated due to its clandestine nature. It is a complex process that includes the exploitation of boys, girls, men and women. Through deception, coercion, physical or psychological threats, or debt bondage, human trafficking victims can become trapped in situations of sexual exploitation, forced labour, servitude, slavery or organ removal.

The Protocol to Prevent, Suppress and Punish Trafficking in Persons, especially Women and Children that supplements the United Nations Transnational Organized Crime Convention came into force in 2000 and defined the crime of human trafficking. Trafficking in persons is not only a grave offence against human dignity, but also ranks among the most lucrative and fastest growing forms of international crime (FAFT and APG, 2018). According to the International Labour Organisation (ILO, 2014), forced labour generates 150.2 billion dollars per year.

A variety of measures to counter this crime can be taken. Counter-trafficking efforts may include addressing risks factors such as lack of employment and income generating opportunities, prevention through awareness raising campaigns, and the facilitation of safe and regular migration channels. Furthermore, ensuring proactive investigation and prosecution of traffickers as well as effective enforcement of protective labour regulations are fundamental for the fight against trafficking (IOM, 2017a, p. 6).

To support prevention and protection activities in human trafficking, an in-depth understanding of victim characteristics is essential. It would allow for timely identification of potential victims of human trafficking (Aronowitz, 2009; Galos et al., 2017).

The analyses from this report – based on IOM's Global Database on Victims of Human Trafficking - aim to get a more nuanced picture of the dynamics at work. They detect the most distinctive victim characteristics related to their entry into the human trafficking process, and the type of exploitation they are subjected to. The analysis focuses on the likelihood of a victim, given a set of (pre-trafficking) characteristics, to be subjected to either forced labour or sexual exploitation. In other words: which victim characteristics predict specific types of exploitation?

They are based on the application of less used methods of analysis in the case of trafficking data - machine learning techniques like clustering and classification tree analysis - which allow to follow a highly inductive approach. These methods lead to new insights emerging directly from the data, rather than letting the empirical analysis to be guided by pre-defined theoretical assumptions. With the application of machine learning techniques complex relationships in the data can be uncovered, which is especially valuable for large datasets such as that analysed.

Current literature on victim characteristics mainly draws on categories like age, sex, country of origin and intended destination (UNODC, 2018; UNODC, 2006). The main findings corroborate some conventional assumptions and illustrate some less described patterns of the way identified victim characteristics and ways of entry into the trafficking process relate to the type of exploitation. Prediction models reveal different dynamics across different regions where IOM is assisting victims of human trafficking. The findings are especially relevant for designing counter-trafficking interventions in the countries of origin.

Findings from the cluster analysis induce to group the data into three main victim types for further analysis. The first group consists of young men and women who are subjected to forced labour in South-Eastern Europe, Asia and the Pacific, and the Middle East and North Africa. The second group consists of young and single women who are sexually exploited in Europe and Central Asia. The third subgroup is characterized by transregional exploitation of young women.

The findings of the classification tree analyses illustrate that there are regional differences regarding the experiences of the identified victims. In South-Eastern Europe and Central Asia, sex and marital status are defining factors in determining the type of exploitation. Citizens of countries from Asia and the Pacific, the predominant form of exploitation is forced labour, regardless of sex. However, women who are recruited through personal contacts and are exploited in the EU and EEA, West and Central Africa or Southern Africa tend to be subjected to sexual exploitation. For citizens of countries from West and Central Africa, the type of exploitation seems to depend on the region where it takes place. While victims exploited in West and Central Africa who are from within the same region are likely to be trafficked for labour exploitation, victims who are trafficked to a continent other than Africa are predicted to be sexually exploited.

Key Findings Analysing Human Trafficking Data – A Machine Learning Approach

The analysis conducted unveiled certain patterns among the identified victims of human trafficking. Certain factors were identified to predict specific types of exploitation. Regional differences regarding the victim characteristics exist. The exploratory nature of the methods of analysis used lead to regional differences in the appearance of dominant characteristic which can explain the patterns of recruitment and exploitation.

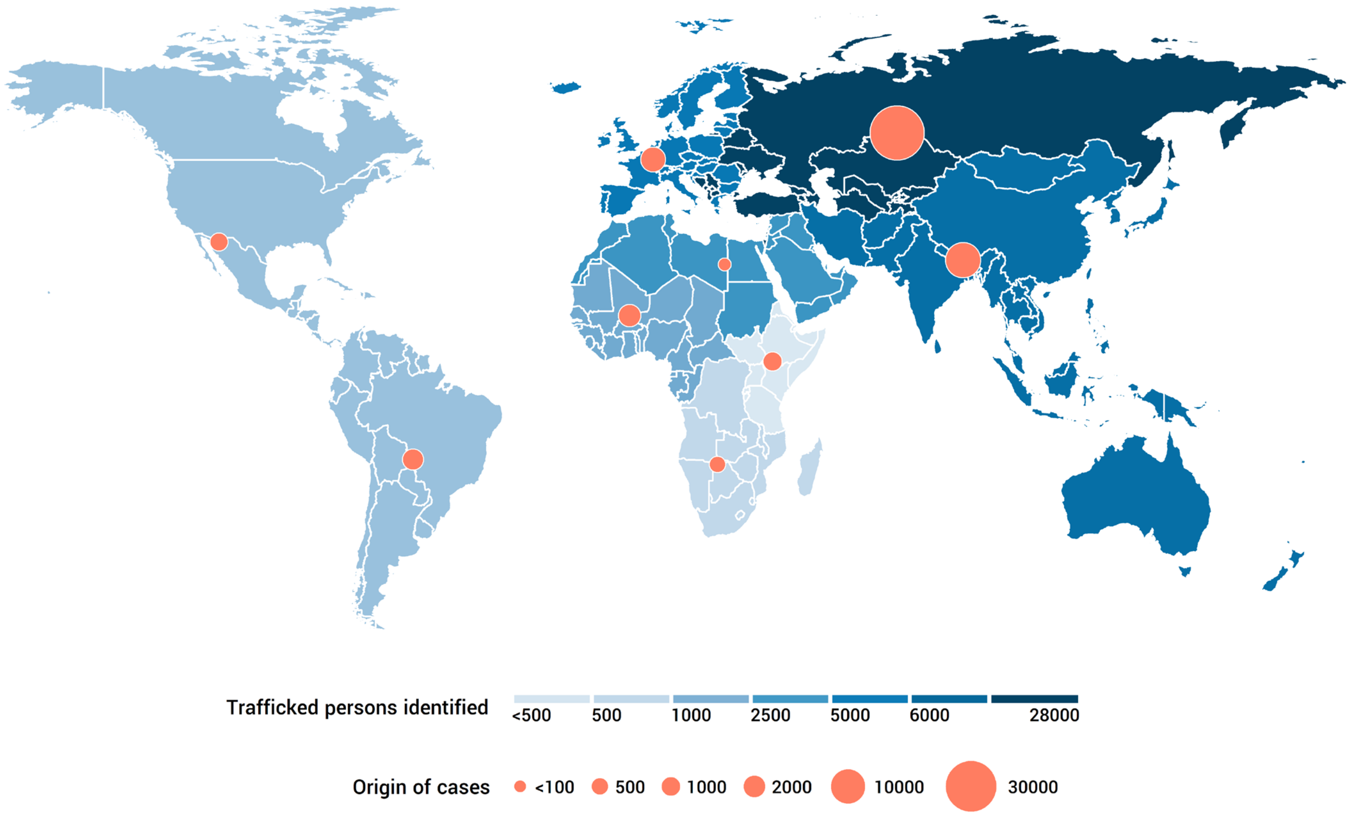
1. The initial contact with the recruiter is a defining experience of victims from West and Central Africa and to some extent of those from the Asia and the Pacific region. Nevertheless, this was not the case for the identified victims from South-Eastern Europe and Central Asia.

2. Women who are nationals of a country from South-Eastern Europe or Central Asia, who married or widowed, tend to be recruited for forced labour. Women from the same regions who are in a partnership other than marriage or who are divorced, separated or single, are typically trafficked for sexual exploitation.

3. In order to predict exploitation type for victims who are citizens from countries in Asia and the Pacific region, age, economic status, and region of exploitation are more influential factors than sex.

4. In the Asia and the Pacific region, identified women who are older than 38 years and have been recruited through a person previously known to them, tend to be subjected to sexual exploitation.

5. Citizens of countries in West and Central Africa tend to be subjected to forced labour when they are being exploited in the region where their country is from. However, the probability of being sexually exploited during the human trafficking process is higher when the victrim is being exploited in another region.



What is Machine Learning?

Machine learning is the practice of using algorithms (a set of rules to solve a problem) to semi-automatically learn from data. There are two main types of machine learning algorithms which differ primarily in the type of task that they are intended to solve. In unsupervised learning, the task is to uncover functions, groupings, and patterns in the data. Therefore, there is no predefined outcome. In supervised learning, a predefined outcome is predicted based on input information (set of features). The goal is to formulate a function that best predicts a certain outcome.

THE DATA

The individual level data analyzed in this report is part of IOM's Global Database on Victims of Human Trafficking. Through its case management activities, IOM has developed the largest database of victims of trafficking case data in the world, with information on nearly 60,000 individual cases.

Disaggregated case-level data are the most detailed source of information on human trafficking and should thus play a vital part of any meaningful analysis on the phenomenon, with due attention to privacy and security concerns. While these types of data have numerous limitations, they are indispensable, as they provide detailed insight into the profiles and experiences of the victims, the forms of human trafficking, and information on perpetrators.

The sample analyzed had over 52,000 cases of identified victims of human trafficking, recorded by IOM from 2002 to 2017. The data used for the analyses presented in this report was very detailed, and the dataset contained numerous fields with unstructured information and notes from caseworkers.

An updated, anonymized and de-identified version of IOM's case data is also publicly available on the Counter-trafficking Data Collaborative (CTDC).

Figure 1 depicts the number of cases registered by IOM, by region of exploitation and region of citizenship. Regions are defined according to the structure of IOM regional offices. The geographic distribution of trafficking victims indicates that human trafficking affects all parts of the world. However, differences across regions exist. The number of identified victims at the time of the analysis ranged from 487 in East Africa and the Horn of

Africa to 28,906 in South-Eastern Europe, followed by the Asia and Pacific region (5,666 identified cases), the European Union and European Economic Area (EU and EEA, 5,236 cases), and the Middle East and Northern Africa (MENA) region (2,934 cases). A vast majority of identified victims originated from South-Eastern Europe and Central Asia (33,972), followed by the Asia and Pacific region (10,026).

The data shows that most exploitation occurs within the same region as the region of origin (81%) and involves transnational exploitation (58%). Out of the identified victims who were exploited trans-regional, the two largest groups are citizens from South-Eastern Europe and Central Asia who are exploited in the EU and EEA, and citizens from Asia and the Pacific who were subjected to exploitation in the MENA region.

Counter Trafficking Data Collaborative

Since the mid-nineties, IOM has assisted over 100,000 victims of trafficking. Through these direct assistance activities, IOM has developed its central case management database which contains information on over 55,000 individual cases since 2002. These data include detailed information on victims of trafficking, including demographics but also information on their trafficking experience. As a unique source of information on human trafficking, IOM has worked to bring these data to a public audience so that valuable insights can be developed and shared among counter-trafficking actors worldwide. A major part of this effort has been the launch of the Counter-Trafficking Data Collaborative (CTDC) in 2017, in partnership with Polaris and Liberty Shared. CTDC is the first global data hub on human trafficking and combines the three largest case-level datasets, resulting in one centralized dataset with information on over 90,000 cases.

https://www.ctdatacollaborative.org/

A GENERAL TYPOLOGY OF VICTIMS OF HUMAN TRAFFICKING

What kind of patterns can be derived from the data? A first and very general typology of the victims of human trafficking was created by an exploratory cluster analysis. The main purpose of clustering is to build meaningful and distinct subgroups, or clusters, that best summarise the data. The three subgroups that emerged can be seen as prototypical profiles of identified victims.

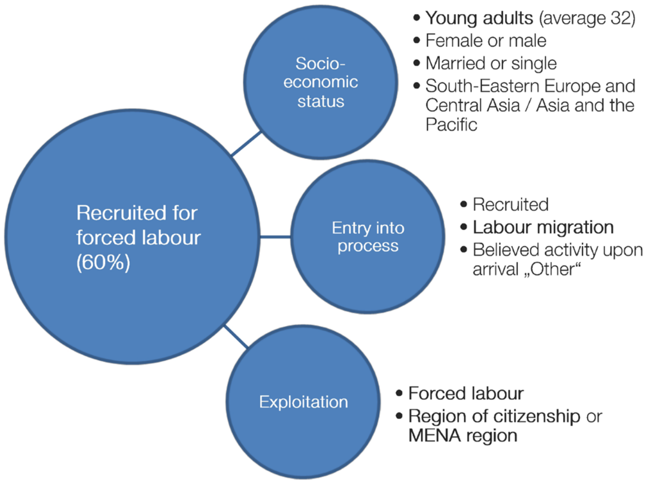


Figure 2. Prototypical characteristics of identified victims of the first subgroup.

Group 1: Recruited for forced labour

The first and largest subgroup represents 60% of the data sample[1] and consists of young adults (on average 32 years old), who are recruited for forced labour (Figure 2). The victims mainly come from South-Eastern Europe and Central Asia, and from Asia and the Pacific region and are exploited in their region of origin, or in the MENA region (around 15%). Marital status is not a very distinctive characteristic of this group, as 44% of the cases were single and 39% married (17% other). The same is true for sex: the male-female ratio is 53% to 47%. This is interesting, as forced labour used to be better known as a form of exploitation for men. With the feminisation of migration however, there are more and more women subjected to labour exploitation, especially when looking at intra-regional movements (Kelly, 2005; Tyner, 1999). A majority of victims belonging to this group entered the trafficking process through labour migration. The literature considers labour migration as a major source of vulnerability to trafficking, as many individuals who migrate might end up having informal working and living arrangements exposing them to exploitation. In that context they cannot turn to the authorities for help, due to irregularity (Clark, 2013). Finally, a striking 71% of the identified victims belonging to this first cluster believed that their activity upon arrival would be "other", meaning an activity that was not pre-defined in IOM's forms and case management system.

Cluster Analysis: Key Findings

- Identified victims of human trafficking in the data can be classified into three distinct subgroups:
  - » The most common victim type represents men and women who are recruited for forced labour and enter the process through labour migration. These are predominantly young adults from South-Eastern Europe and Central Asia, and Asia and the Pacific who are exploited in their region of citizenship and the MENA region.
  - » Another reoccurring victim type encompasses young and single women from South-Eastern Europe and Central Asia who are subjected to sexual exploitation in Europe and Central Asia.
  - » The third group is mostly defined by the transregional character of the exploitation. These are mostly young, single, women from parts of sub-Saharan Africa or South-Eastern Europe and Central Asia exploited in the MENA region or the EU and EEA.
- The findings confirm some general assumptions about patterns of human trafficking, for example "young women tend to be victims of sexual exploitation", or "labour migration renders individuals more vulnerable" (Clark, 2013; UNDOC, 2006).
- Type of exploitation, along with region of citizenship and region of exploitation appear to be determinant factors in grouping the identified victims of human trafficking.

[1] Because Cluster Analysis requires a lot of computational power, the analysis was done on a random sample of 1000 observations. Therefore, 60% of the sample are 600 individuals.

Group 2: Sexual exploitation in Europe and Central Asia

The second subgroup that emerged from the cluster analysis represents identified victims of sexual exploitation in Europe and Central Asia, depicted in Figure 3. In contrast to the first cluster, members of the second cluster are almost exclusively female (98%) and single (72%). Also, they tend to be younger (23 years on average) than the victims described in the first group. There is little information recorded on the entry into process for the second group, except that the majority has been recruited (79%). These women are predominantly citizens of South-Eastern Europe and Central Asia and are mainly exploited in their region of origin, but also in the EU and EEA (25%). This trafficking pattern could be explained by some of the post-Cold War changes. Some consider, the opening of borders between East and West after the fall of the Soviet Union provided new opportunities for criminal actors exploiting the widespread poverty in South and Eastern Europe, resulting in a remarkable increase in human trafficking. Many women sought work abroad and were deceived and subsequently forced into prostitution (Kligman & Limoncelli, 2005; Surtees, 2008).

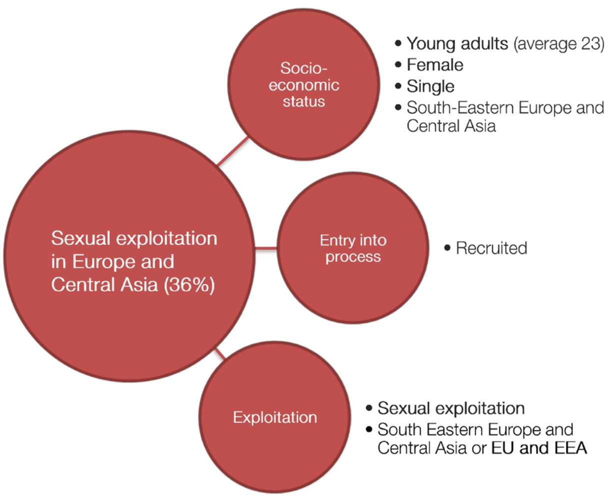


Figure 3. Characteristics of identified victims belonging to Group 2

Group 3: Trans-regional exploitation of young women

The third cluster, seen in Figure 4, is much smaller than the first two, making up only 4 percent of the total sample. What sets this group apart is that the identified victims were not recruited (72%) and the majority was trans-regionally exploited (75%). The victims mainly originate from sub-Saharan Africa (44%) and South-Eastern Europe and Central Asia (36%) and are predominantly young women. Over half of the identified victims assigned to this cluster were exploited in the MENA region, about one third suffered exploitation in the EU and EEA.

Another main difference between this third group compared to the other two is that the exploitation type is not a defining feature. Almost half of the cases belonging to this group were subjected to sexual exploitation (47%). 55% of the identified victims were subjected to an exploitation type "other" than the ones predefined by IOM. In addition, 30% were identified as victims of forced labour[2].

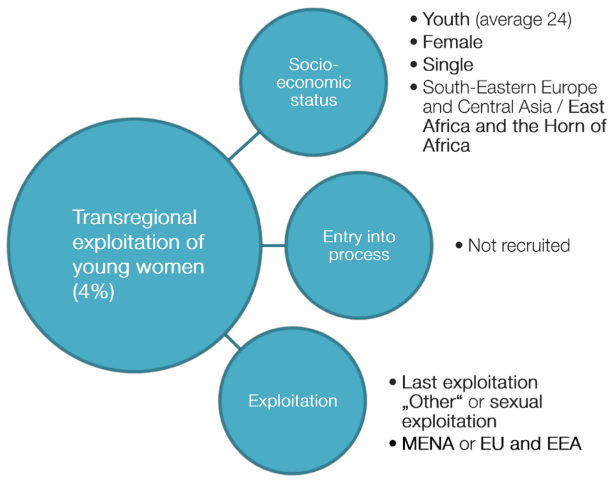


Figure 4. Characteristics of identified victims belonging to Group 3

What is Clustering?

Cluster analysis is an unsupervised machine learning technique used to detect patterns and groupings in the data. Using a predefined way of measuring similarity, observations that are similar to one another are grouped into the same cluster, while observations in different clusters should be clearly distinguishable from one another. More detailed information about the clustering process and how similarity is measured can be found in the "'Data Analysis on Human Trafficking' – Methodology Report".

[2] *An identified victim can report various types of exploitation also from different experiences.*

REGIONAL PATTERNS IN DETERMINING TYPE OF EXPLOITATION

Country of origin, their region of exploitation, and the type of exploitation are the main characteristics that group the identified victims. They are subjected to mainly forced labour or sexual exploitation.

Classification tree models, a form of supervised machine learning and useful to uncover complex interactions, were used to predict the type of exploitation. Given the provided information on victim characteristics, the model learns how these input characteristics (available victim data that fed into the model) play together in predicting a certain outcome. The input factors fed to the algorithm were socio-economic factors, as well as information regarding the entry into the process of trafficking. The outcome sought to predict was whether an identified victim would be subjected to forced labour or sexual exploitation[3].

Based on the selected information regarding socio-economic factors as well as entry into the trafficking process, the algorithm tries to define rules to best classify the identified victim cases according to type of exploitation, resulting in a "tree".

Of all regional tree models that were generated three were chosen for their high predictive power. The first concerns citizens of South-Eastern Europe and Central Asian country. The second tree classifies citizens from Asia and the Pacific countries. The last tree was estimated on a subset containing citizens from West and Central African countries. The three regions were chosen because they are major regions of origin and are good examples in illustrating the regional differences in the patterns determining type of exploitation.

What are Classification Trees?

Classification trees are used to effectively predict a predefined outcome (e.g. exploitation type) based on the values of selected input variables (e.g. socio-economic factors). Starting with the whole sample data, the algorithm chooses the input information that is most insightful (e.g. marital status) in classifying the observations in order to obtain homogeneous subgroups according to the selected outcome (e.g. forced labour or sexual exploitation). This step is repeated, resulting in various "splits" in the data. The results can be visualised as a tree, with the various splits as branches and the subgroups as leaves. This analysis relied on the Gini Index to build classification tree. The Gini Index measures how good a split is by looking at how many observations were misclassified, e.g. how many cases were misclassified as labour exploitation when they were in fact cases of sexual exploitation. At each step, the algorithm chooses the split that maximally reduces the Gini Index. More detailed information about the Gini Index and how the tree models were built can be found in the "Data Analysis on Human Trafficking' – Methodology Report".

[3] *In an earlier stage, the outcome category "both labour and sexual exploitation" was included as well. However, as this category never showed up in the trees, observations subjected to both labour and sexual exploitation were removed from the sample. 575 cases were affected.*

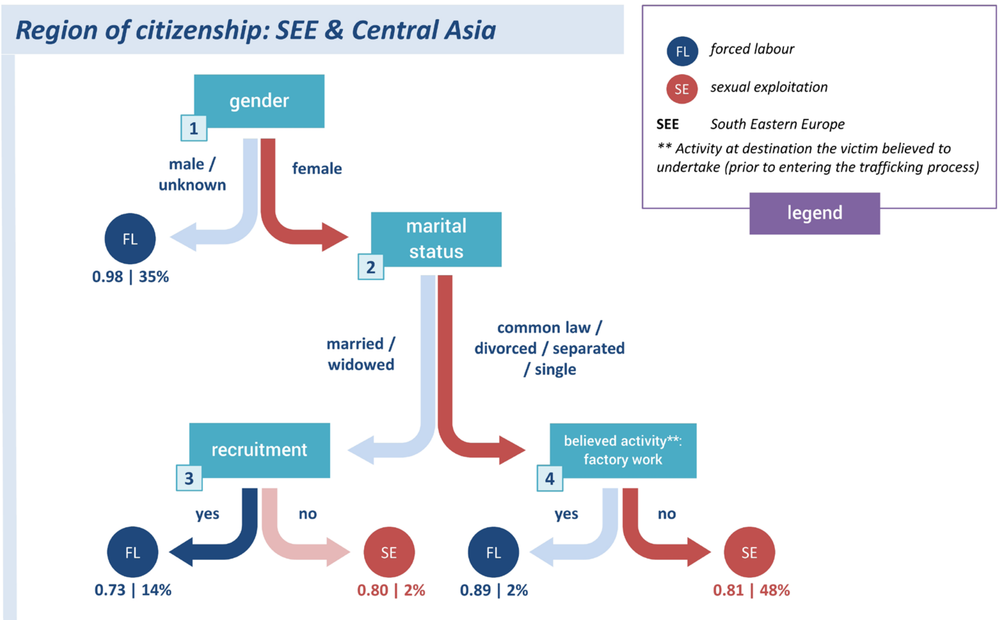


Figure 5. Classification tree predicting type of exploitation for citizens of SEE & CENA

SOUTH-EASTERN EUROPE AND CENTRAL ASIA: SEX AND MARITAL STATUS AS DEFINING FACTORS

As mentioned above, over half of the identified victims in the dataset come from South-Eastern Europe and Central Asia (SEE & CENA). Overall, 33,972 of the identified victims were citizens from South-Eastern European and Central Asian countries, while 28,906 were exploited in this region.

7,696 identified trafficking victims originating from SEE & CENA were subjected to forced labour, and 6,329 were sexually exploited. Given these high numbers, knowing more about how victim characteristics and exploitation type relate in this region is of particular interest.

Figure 5 shows the most important characteristics in determining the type of exploitation for citizens of South-Eastern Europe and Central Asian countries. Sex appears to be a crucial factor. Identified victims who are male or whose sex is not defined have a 98% predicted probability of being a victim of forced labour rather than sexual exploitation. 35% of the observations in the sample follow this pattern, these are around 3,900 cases. Looking at the female victims, marital status is a defining characteristic. For women who are not "married" or "widowed" the predicted probability of being subjected to sexual exploitation is 81%. Only exception are women who reported that they believed prior to entering the process that they will work in factories upon arrival. They were exploited in forced labour (2% of the sample with 86% probability). 48% of the sample (5,404 cases) are classified into this group. By contrast, women who are married or widowed and for whom recruitment is part of the trafficking process tend to be subjected to forced labour (predicted probability of 73%, 1,535 cases).

The model, making predictions about identified victims from South-Eastern Europe and Central Asia, seems to corroborate the findings of the cluster analysis and some general assumptions made in the literature. Being male is a strong predictor for forced labour. In 2016, 95% of all identified male victims of trafficking were trafficked into forced labour IOM (2017b). Also, 85% of the male victims in the dataset confirmed the question whether they had been victims of labour exploitation, while under 3% of the identified male victims reported to have been subjected to sexual exploitation. As men might fear to speak up about their own sexual exploitation more than women do, due to higher stigma or perceptions on masculinity, the number of unreported cases might be considerably higher. Another major problem in identifying male victims of human trafficking is legislation. Laws relevant for counter-trafficking are restricted in their application to women in many countries, in some cases only to women who are trafficked for sexual exploitation (UNDOC, 2006, p.78).

Other than sex, marital status reappears as an influential factor in determining type of exploitation, but only for women. According to the literature dealing with human trafficking in South-Eastern Europe and Central Asia, women were especially hit by unemployment after the breakdown of the Soviet Union. In addition to these economic considerations, reasons for migration and trafficking often included factors like domestic violence or stigmatization of female-headed households (Kligman & Limoncelli, 2005; Surtees, 2008; Langberg, 2005). The story of one of these women is illustrated in the quote above. The quote was taken from information collected by IOM case workers who conducted interviews with assisted victims of human trafficking.

In general, the female profile depicted in Figure 5 corresponds to the findings found in Group 2 of the cluster analysis: single women tend to be pulled into sexual exploitation. However, there are also cases that counter this general assumption. Women whose marital status is either "common law", "divorced", "separated" or "single" have a much higher predicted probability of falling into forced labour rather than sexual exploitation, if they believe that they are going to be doing factory work in the intended destination country[4]. Around 200 cases fall into this category. Also, while married or widowed women tend to be trafficked for labour exploitation, this pattern seems to depend on whether a woman experiences recruitment as part of the trafficking process. Women who are married or widowed and not recruited have a predicted probability of 80% of being subjected to sexual exploitation, rather than labour exploitation.

"After the early death of her parents the life of Ms W was very difficult. Her ex-husband did not help her and her child at all.

The amount she earned working at the local market was not enough. Ms W was in need and was searching for a better paid job or for an opportunity to work abroad for some time. [...]"

Anonymised quote obtained from interviews with identified victims of human trafficking by IOM caseworkers.

[4] *Identified victims reporting labour migration were asked for their intended destination and the activity they will be doing after arriving at the intended destination.*

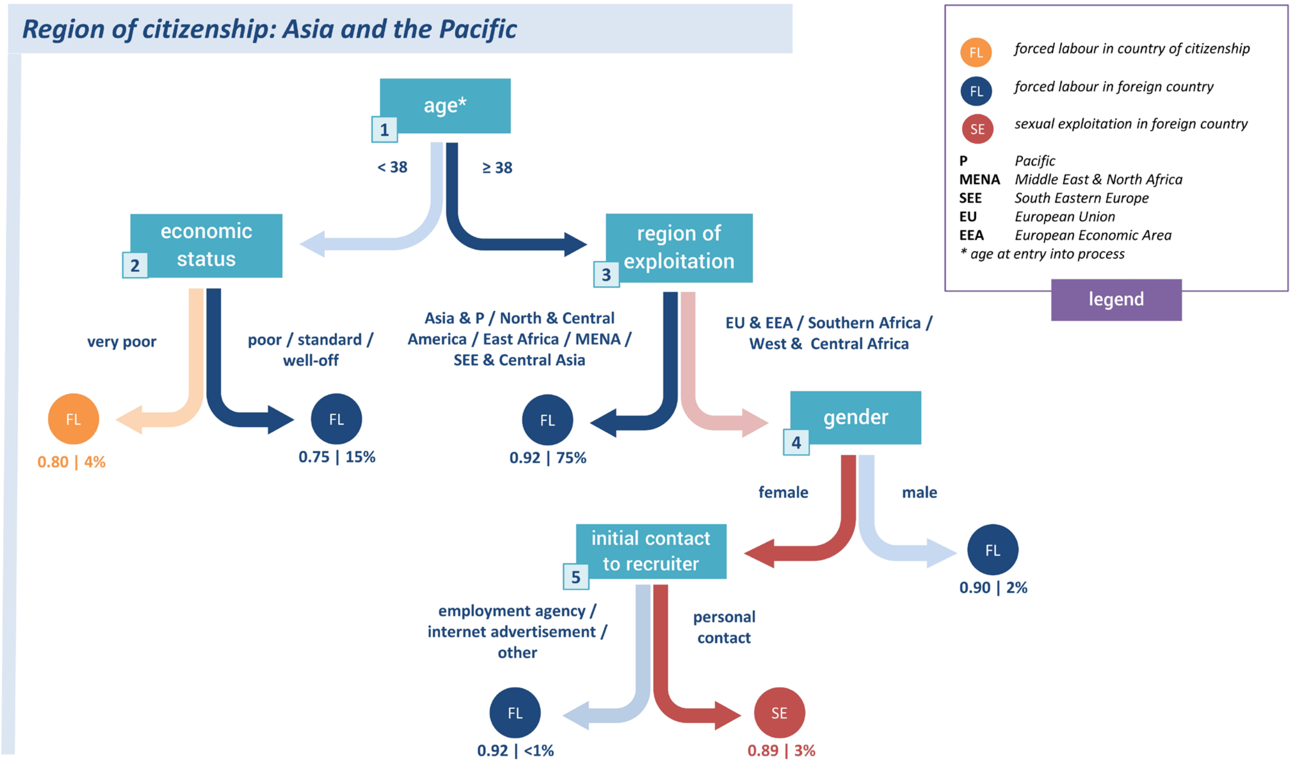


Figure 6. Classification tree predicting type of exploitation for citizens of Asia and the Pacific

How to read the Tree?

The classification tree can be read in a similar way as a flowchart.

Each numbered box stands for a split in the data based on a specific attribute. The circles represent the outcome: either sexual exploitation, forced labour, or, in some cases, forced labour in the country of citizenship, and forced labour in a foreign country.

The numbers beneath the circles are the predicted probabilities (on the left), and the percentage of cases that fall into a given outcome category (on the right).

The predicted probability is a measure of how many individuals are expected to fall into the same outcome category, given that they share the same characteristics.

The higher up an attribute / characteristic appears in the tree, the more insightful it is in classifying the observations based on the outcome variable.

For example, Figure 5: Victims, who are female, married or widowed, and are recruited, have a 0.73 (73%) predicted probability of landing in labour exploitation compared to sexual exploitation. 14% of the observations included are represented with this rule.

ASIA AND THE PACIFIC: FORCED LABOUR AS PREDOMINANT TYPE OF EXPLOITATION

Asia and the Pacific constitutes the second largest region of origin in the dataset. Citizens of this region are most dominantly subjected to forced labour. Over 5,700 individuals in this data were identified as victims of labour exploitation, while under 500 cases were registered as victims of sexual exploitation. For over one third of the cases originating from Asia and Pacific, the information regarding the type of exploitation was not recorded by the case workers

Figure 6 depicts the classification tree for the citizens of Asia and the Pacific. Age and perceived economic status seem to play together in deciding where exploitation takes place. While victims in Asia and the Pacific generally tend to be subjected to forced labour in a foreign country, identified victims who are younger than 38 years old and describe their economic status as very poor tend to be exploited within the borders of the home country. This pattern is expected to hold 80 out of 100 times and is based on 4% of the observations (around 200 cases). Compared to these, individuals who are younger than 38 years old but describe their economic status as either poor, standard, or well-off rather than very poor, have a 75% predicted probability of being trafficked for forced labour across national borders (over 600 cases). Individuals who are 38 years old or older are predicted to be trafficked for forced labour across borders, if the region of exploitation includes Asia and the Pacific, North and Central America, East Africa, the MENA region, or South-Eastern Europe and Central Asia. A large majority of 75% of the observations (more than 3,000 cases) fall into this pattern. Note that sex is not a defining factor regarding the exploitation type for observations falling into split 1-3. Sex becomes relevant only at split number 4: Citizens from Asia and the Pacific, who are 38 years old or older, exploited in the EU and EEA, Southern Africa, or West and Central Africa, and are men, are predicted to be victims of forced labour in a foreign country (predicted probability of 90%, around 80 cases). By contrast, women who share the same characteristics tend to be subjected to sexual exploitation in a foreign country, depending on the initial contact to the recruiter. Women who are at least 38 years old, exploited in the EU and EEA, Southern Africa, or West and Central Africa are more likely to be sexually exploited, if contact to the recruiter was initiated through personal contact (predicted probability of 89%, around 150 cases). Compared to that, for women with the same characteristics, recruitment techniques like employment agencies, internet advertisement and "other" seem to be more closely related to labour exploitation in a foreign country (predicted probability of 92%). However, less than 20 cases are represented by this rule.

What stands out in the classification tree depicted in Figure 6 is the rather high age of the women who become victims of sexual exploitation. Since there is the predominant idea that women who suffer sexual exploitation are young, or even minors (Kelly, 2005), the pattern in the findings concerning women who are 38 years old or older runs against usual expectations. One possible explanation is that these women might already work in the sex-related industry and find themselves in difficult financial situations. They might be aware that they are being transported to regions like Europe to work in the sex industry. They are, however, unaware of the slave-like working conditions they end up in (Melrose & Barrett, 2006; Vijeyarasa, 2012).

Often, they are recruited through personal contact and in many cases by other women, sometimes former victims of trafficking. The recruiters look for vulnerable women within family networks and/or ethnic communities (Europol, 2011). The story of an identified victim with such a profile can be illustrated by the quote below.

"Ms KN was approached by the recruiter in a pub in Thailand. In this pub Ms KN was working as a waitress and as a dancer. The recruiter promised her the same job in Switzerland but promised her a higher salary, which would allow Ms KN to take care of her family. Due to the poor economic situation of her family, Ms KN accepted the job offer and the recruiter organised everything (flight and visa) for her against a commission (approx. CHF 30,000) which she was told she can pay back once in Switzerland. When Ms KN arrived in Switzerland, she was immediately brought to a club, where she had to prostitute herself in order to repay the debts [...]"

Anonymised quote obtained from interviews with identified victims of human trafficking by IOM caseworkers.

"She was recruited by a 'madam' in Benin City while she was working as a hairdresser with her cousin. The woman offered them to work in Italy as hairdressers to earn the amount of EUR 400/month. She travelled with her cousin all the way to Europe through Niger and Libya (where they spent about a month waiting to embark on a boat). Once arrived in Italy, they contacted another 'madam' who told themthat there was no job as hairdresser but instead in prostitution. They were forced to provide sexual services to clients, could not refuse any clients. When they refused or complained, they were beaten or threatened. The little money they got, they had to give it to the 'madame' and her husband (approx. 100 EUR). One day, they managed to escape […]."

Anonymised quote obtained from interviews with identified victims of human trafficking by IOM caseworkers.

WEST AND CENTRAL AFRICA: DIFFERENT DESTINIES DEPENDING ON REGION OF EXPLOITATION

With 1,889 identified victims originating from this region, West and Central Africa constitutes the fourth largest region of origin in this data. 694 cases were subjected to labour exploitation, while 284 were trafficked for sexual exploitation. For over 900 individuals, information regarding the type of exploitation is missing.

Looking at Figure 7, one can see that the region of exploitation plays an influential role in defining the type of exploitation a victim from West and Central Africa is likely to be subjected to. In general, it seems that victims from this region are predicted to be victims of forced labour, if they are trafficked for exploitation on the African continent. By contrast, victims who are transported to another continent tend to land in sexual exploitation. A pattern that catches one's eye is the path highlighted in orange leading from split number 1 to 2. If the identified victims are trafficked within West and Central Africa and are sold by members of their family, they are predicted to be exploited in forced labour within the borders of their country of citizenship (with a 98% predicted probability). Around 450 observations are classified into this subgroup. For victims who are trafficked within the West and Central African region but initiated contact to the recruiter through personal contact or other means, rather than being sold by family members, the predicted type of exploitation is forced labour in a foreign country (around 70 cases). However, the predicted probability associated with this rule is not very high (62%).

Child trafficking is a prevalent problem in the region of West and Central African. UNICEF (2006) estimates that hundreds of thousands of children are trafficked throughout the region. Often, these children are sold by their parents or other relatives. Driven by extreme poverty, persistent unemployment, armed conflicts, and human deprivation, these parents often have no choice but give their children away (Adepoju, 2005). They hope that their children will be looked after in a better place, with better opportunities and the payment they receive is rationalised as an advance on their children's salaries (Bales, 2007, p. 269). Traffickers promise to treat the children well or provide them with an education. In many circumstances, however, they are exploited for agricultural and domestic labour, or begging, and forced to work under slave-like working conditions. Often, they are exposed to psychological, physical, and sexual abuse (Adepoju, 2005; ILO, 2003, Sawadogo, 2012).

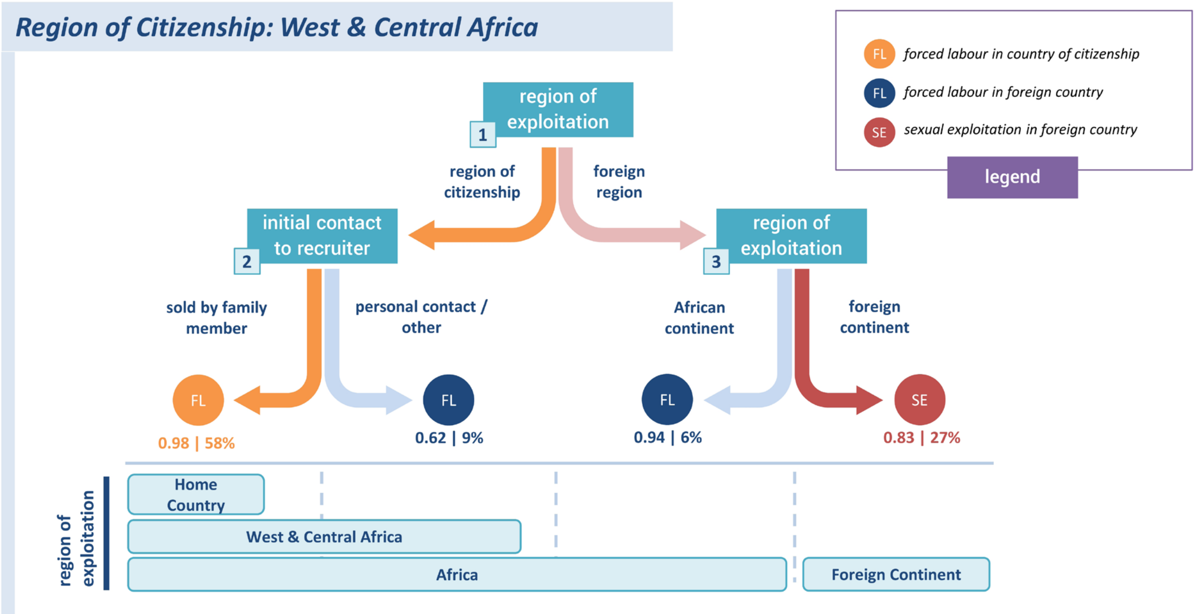


Figure 7. Classification tree predicting type of exploitation for citizens of West and Central African countries

Moving to the right-hand side of the classification tree depicted in Figure 7, one can see that victims of West and Central Africa who are trafficked onto another continent than Africa tend to end up in sexual exploitation, with a predicted probability of 83% (around 220 cases), while victims exploited on the African continent tend to be trafficked for forced labour (predicted probability of 94%, around 50 cases). Transnational trafficking in human beings has been a serious problem for years in the region of West Africa and Central Africa but has taken new complex and dramatic dimensions in the wake of globalization. Facilitated by the porosity of regional and continental boundaries, and governments weakened by corruption, girls and young women are transported and forced into prostitution in regions like Europe, the United States, and the Gulf states (Adepoju, 2005; Mazzitelli, 2007; Sawadogo, 2012).

Well organised criminal networks take charge of the victim from the moment of recruitment to his or her exploitation (Mazzitelli, 2007; Sawadogo, 2012). According to Mazzitelli (2007, p. 1079), a recruiter of a woman trafficked for sexual exploitation spends approximately USD 2,000 to bribe the right officials, obtain travel documents and safe houses, and transport the woman to a "madam". This madam then pays approximately USD 12,000 for the woman. Often, victims are resold as they approach the term of their contracts. An exemplary story for such a destiny is illustrated by the quote above.

Classification Tree Analysis: Key Findings

- Country of origin, their region of exploitation, and the type of exploitation are the main characteristics that group the identified victims. They are subjected to mainly forced labour or sexual exploitation.
- Identified victims from South-Eastern Europe and Central Asia who are male or whose sex is not defined have a 98% predicted probability of being a victim of forced labour rather than sexual exploitation. 35% of the observations in the sample (3,900 cases) follow this pattern.
- While victims in Asia and the Pacific generally tend to be subjected to forced labour in a foreign country, identified victims who are younger than 38 years old and describe their economic status as very poor tend to be exploited within the borders of their home country.
- Identified victims from West and Central Africa, who are trafficked for exploitation on the African continent, are predicted to become victims of forced labour. If they are transported to another continent they tend to end in sexual exploitation. But if they are trafficked within West and Central Africa and are sold by members of their family, they are predicted to be exploited in forced labour within the borders of their country of citizenship (with a 98% predicted probability).

LIMITATIONS

The findings of this analysis must be understood in the light of limitations of data collection, and of the methods of analysis.

The most important aspect related to the limitations of this type of data, is to acknowledge that it is operational data that is a reflection of identification and assistance capacities, in particular environments. The dataset on identified victims of human trafficking is limited in geographic scope and comprehensiveness. For example, IOM country offices' capacity to collect data on identified victims vary between countries. Not all regions are represented equally in terms of the number of identified victims. This selection is based on resource constraints, security reasons, or presence of large-scale humanitarian emergencies. This means that data are only available where such organizations are operational and able to share such data. In practice, data are therefore not available for all countries/locations and, where data do exist, they are not always comprehensive in terms of coverage of a given country/location.

In addition, large quantities of data on identified victims of trafficking do not necessarily indicate higher prevalence of human trafficking. Indeed, they may equally indicate an effective counter-trafficking response. Identified cases are better understood as a sample of the unidentified population of victims, yielding insight into trafficking trends and patterns. This sample may be biased if some types of trafficking cases are more likely to be identified (or referred) than others. The extent of the bias is not always known or able to be corrected for, since the unidentified population is, by definition, unknown

In terms of geographical representation of victims, data on victims is almost absent from humanitarian emergencies. For example, reaching out to victims of trafficking in the context of crisis for example, is difficult as security may impinge access to this affected populations, or migrants may be caught in an irregular situation, without access to legal documentation. They may fear or have been threatened with arrest, which makes them unwilling/unable to share relevant information. Further, the forms of exploitation related to trafficking in person – whether sexual, slavery or labour exploitation – make it particularly challenging to interview victims, who may fear further stigmatization and social marginalisation. Therefore, different levels of trust and different cultural norms influence the willingness of a victim to participate in the survey. It is very sensitive to talk about experiences of exploitation and trafficking that could be traumatic (Aronowitz, 2009; Galos et al., 2017).

Another limitation of the results is the accuracy of answers from respondents, which in most cases cannot be verified. Self-reported nationalities, for instance, cannot be ascertained beyond doubt, and could be a source of error.

The methodological approach has also limitations. The hierarchical clustering was only performed with a random sample of a thousand observations due to computational constraints. This limits the representativeness and stability of the findings. Further, decision tree models are to some degree unstable, especially for regions with few identified victims. One solution that was applied was to carry out the analysis for slightly different versions of the dataset and select decision tree models that are sufficiently robust. For the decision tree models based on well-covered regions with a large sample in the dataset a higher robustness could be detected. Along with the instability, decision tree models are faced with a complexity problem with a tendency to generate overfitted models. In other words, the attempt to make the model adapt too closely to slightly inaccurate data can infect it with considerable errors and diminish its predictive power. This is not desirable as the complexity problem mentioned above hinders the generalization of decision tree models across the population. The findings must be understood as an optimal trade-off between accuracy of the models and the complexity of the models.

CONCLUSION

As the data illustrate, human trafficking affects all regions worldwide.

The cluster analysis resulted in a first and very general typology of identified victims of human trafficking including three prototypical victim types. The first and largest subgroup consists of young men and women who are subjected to forced labour and enter the trafficking process through labour migration. These victims originate mainly from South-Eastern Europe or Asia and the Pacific. They are exploited in their region of origin or the Middle East and North Africa. The second group consists of young and single women who are sexually exploited in the EU and EEA or South-Eastern Europe and Central Asia. Most victims categorized in this group are citizens of South-Eastern Europe and Central Asia. The third and smallest subgroup is characterized by transregional exploitation of young women. These women are predominantly from parts of sub-Saharan Africa or South-Eastern Europe and Central Asia. They are trafficked for exploitation in the MENA region or the EU and EEA.

The classification tree analysis with a focus on type of exploitation provides a more nuanced picture of the dynamics at work. The aim was to understand which victim characteristics best predict whether an individual is going to be subjected to forced labour or sexual exploitation.

The findings of the classification tree analysis illustrate that regional differences regarding the prevalence of victim characteristics exist. Looking at citizens from South-Eastern Europe and Central Asia, sex proved to be a strong predictor of type of exploitation. Men and victims whose sex is unknown tended to be subjected to forced labour rather than sexual exploitation. For women originating from South-Eastern Europe and Central Asia, marital status could be seen as an indicator of the type of exploitation. Married or widowed women tended to be recruited for forced labour with an exception of the ones who were not recruited, they were mainly sexually exploited. Based on the classification tree model women whose marital status is common law, divorced, separated or single are expected to be trafficked for sexual exploitation. Women who believed that they were going to work in a factory were the most likely to be exploited for forced labour.

In comparison, citizens from Asia and the Pacific countries were predominantly exploited for forced labour in a foreign country, regardless of their sex. An exception to this pattern were women who were older than 37 years, and who were recruited through personal contact and exploited in the EU and EEA, Southern Africa or West and Central Africa. These women tended to be subject to sexual exploitation. Women who shared the same characteristics but were recruited by an employment agency, internet advertisement or other, are more likely to be recruited for labour exploitation. Furthermore, victims who are younger than 38 years and described their economic status as very poor, rather than poor, standard or well-off, are predicted to be subject to forced labour in their home country rather than in a foreign country.

Finally, identified victims originating from West and Central Africa tended to be trafficked mostly for forced labour. The region of exploitation seems to be a crucial factor in identifying whether citizens from this region experience forced labour or sexual exploitation. Victims who were transported to a continent other than Africa are predicted to fall into sexual exploitation, while victims who were trafficked in Africa are more likely to suffer labour exploitation. Within the region of West and Central Africa, citizens tended to be subject to forced labour in a foreign country if they were recruited through personal contacts or other types of recruitment. By contrast, victims who were sold by members of their family tended to be forced into labour exploitation within the borders of their country of citizenship. These were mostly children who were given away by their parents due to the extremely dire circumstances at home.

The analysis confirms the prevalence of dominant victim characteristics discussed in the literature such as age, sex, origin, and an individual's economic status. In addition, the results emphasize other relevant characteristics like marital status, the activity a victim believes to be undertaking in the intended destination country, and the initial contact to the recruiter.

Most importantly, the classification tree models allowed the assessment of the relative importance of a characteristic among a wide range of characteristics.

Besides complementing the understanding of victim characteristics and how they relate to type of exploitation, the analysis shows how the application of advanced analysis techniques can be advantageous in the field of human trafficking research. The inductive approach enables the identification of victim characteristics moving beyond pre-assumed social groups. Further, the machine learning strategies allowed to map complex interactions of such characteristics. The approximated models were directly generated through a learning process based on the data. Classification tree analysis has the advantage of dealing well with datasets containing many missing values. In the field of human trafficking research with often very limited access to information, this might be of special interest.

Consequently, machine learning techniques showed that structured quantitative empirical analysis has much potential to support counter-trafficking efforts. However, the importance of extensive, high quality, and standardized data in assessing the explanatory power and validity of empirical findings concerning human trafficking cannot be stressed enough. Nevertheless, these results must be considered in the light of their limitations. The data includes victims identified and assisted by IOM only, which constitutes a strong selection bias. The precise extent of trafficking remains unknown and there are countless numbers of trafficking victims who are never identified (IOM, 2017b). Keeping these limitations in mind, the results of the analysis are still based on thousands of real victims representing real stories of a high number of men, women, and children around the world. Even though not generalizable to all of the trafficking victims, understanding the trends and characteristics within a subpopulation is still essential for counter-trafficking efforts.

Applying innovative research methods is an imperative in complementing existing research on the trafficking in persons (Andrees, & Van der Linden, 2005). Combining advanced statistical techniques with an elaborated data collection strategy set the starting point for evidence based, reliable counter-trafficking solutions, their implementation and the monitoring of the policy effects.

REFERENCES

Adepoju, A. (2005). Review of research and data on human trafficking in sub-Saharan Africa. *International Migration*, 4 (1-2), 75-98.

Andrees, B., & Van der Linden, M. N. (2005). Designing 1Trafficking Research from a Labour Market Perspective: The ILO Experience1. *International Migration*, 43(1-2), 55-73.

Aronowitz, Alexis. (2009). Guidelines for the Collection of Data on Trafficking in Human Beings, Including Comparable Indicators. (Vienna: Federal Ministry of the Interior Austria & International Organization for Migration (IOM)). Retreived from http://publications.iom.int/system/files/pdf/guidelines_collection_data_iomvienna.pdf

Bales, K. (2007). What Predicts Human Trafficking?. *International Journal of Comparative and Applied Criminal Justice, 31*(2), 269-279.

Clark N. (2013). *Detecting and Tackling Forced Labour in Europe* (Joseph Rowntree Foundation Report). Retrieved from https://www.jrf.org.uk/report/detecting-and-tackling-forced-labour-europe

Europol. (2011). Knowledge Product: Trafficking in Human Beings in the European Union (08 OC Networks in the South-East European Sphere / 02 Analysis and Knowledge / September 2011). Retrieved from https://www.europol.europa.eu/node/468

Financial Action Task Force (FATF) and Asia/Pacific Group on Money Laundring (APG). (2018). Financial Flows from Human Trafficking. (Paris.) Retrieved from https://www.fatf-gafi.org/media/fatf/content/images/Human-Trafficking-2018.pdf

Galos, E., Bartolini, L., Cook, H. & Grant, N. (2017). Migrant Vulnerability to Human Trafficking and Exploitation: Evidence from the Central and Eastern Mediterranean Migration Routes. (Geneva: International Organization for Migration). Retrieved from https://publications.iom.int/system/files/pdf/migrant_vulnerability_to_human_trafficking_and_exploitation.pdf.

International Labour Organization. (2003). Trafficking in Human Beings. New Approaches to Combating the Problem. Retrieved from https://www.refworld.org/docid/40360bf34.html.

International Labour Organization. (2014). Profits and poverty: The economics of forced labour. Retrieved from https://www.ilo.org/wcmsp5/groups/public/---ed_norm/---declaration/documents/publication/wcms_243391.pdf

International Organization for Migration. (2017a). Combating Trafficking in Persons and Contemporary Forms of Slavery. Retrieved from https://www.iom.int/sites/default/files/our_work/ODG/GCM/IOM-Thematic-Paper-Trafficking-in-persons.pdf

International Organization for Migration. (2017b). Global Trafficking Trends in Focus IOM Victim of Trafficking Data 2006 – 2016. (August 2017). Retrieved from https://www.iom.int/sites/default/files/our_work/DMM/MAD/A4-Trafficking-External-Brief.pdf

Kelly L. (2005). 'You Can Find Anything You Want': A Critical Reflection on Research on Trafficking in Persons within and into Europe. *International Migration*, 41, 235-263.

Kligman, G., & Limoncelli, S. (2005). Trafficking women after socialism: From, to, and through Eastern Europe. Social Politics: *International Studies in Gender, State and Society, 12*(1), 118-140.

Langberg, L. (2005). A review of recent OAS research on human trafficking in the Latin American and Caribbean region. *International Migration*, 43(1-2), 129-139.

Mazzitelli, A. L. (2007). Transnational organized crime in West Africa: the additional challenge. *International Affairs*, 83(6), 1071-1090.

Melrose M. & Barrett D. (2006). The Flesh Trade in Europe: Trafficking in Women and Children for the Purpose of Commercial Sexual Exploitation. *Police Practice and Research, 7*(2), 111-123.

Sawadogo, W. R. (2012). The challenges of transnational human trafficking in West Africa. *African Studies Quarterly, 13*(1/2), 95.

Surtees, R. (2008). Traffickers and trafficking in Southern and Eastern Europe: Considering the other side of human trafficking. *European Journal of Criminology, 5*(1), 39-68.

Tyner J.A. (1999). The Web-based Recruitment of Female Foreign Domestic Workers in Asia. *Singapore Journal of Tropical Geography, 2*(2), 193-209.

United Nations International Children's Emergency Fund (UNICEF). (2006). West and Central African nations join forces to end child trafficking. Retrieved from https://www.unicef.org/protection/nigeria_34868.html

United Nations Office on Drugs and Crime (UNODC). (2006). Trafficking in Persons: Global Patterns. Vienna. Retrieved from http://www.unodc.org/documents/human-trafficking/HT-globalpatterns-en.pdf

United Nations Office on Drugs and Crime (UNODC). (2008). An Introduction to Human Trafficking: Vulnerability, Impact and Action. Vienna. Retrieved from http://www.unodc.org/documents/human-trafficking/An_Introduction_to_Human_Trafficking_-_Background_Paper.pdf

United Nations Office on Drug and Crime (UNODC). (2016). Global Report on Trafficking in Persons (United Nations Publication / Sales No. E.16.IV.6.) Retrieved from https://www.unodc.org/documents/data-andanalysis/glotip/2016_Global_Report_on_Trafficking_in_Persons.pdf

Vijeyarasa R. (2012). The Cinderella syndrome: Economicexpectations, false hopes and the exploitation of trafficked Ukrainian women. *Women's Studies International Forum*, 35, 53-62.

APPENDIX

Table 1. List of countries coded as regions (according to IOM; additional changes were made for statistical reasons)

EU and EEA		South-Eastern Europe and Central Asia	Middle East and North Africa	Asia and the Pacific		West and Central Africa	East Africa and the Horn of Africa	Southern Africa	South America	Central America, North America and the Caribbean
Andorra	Slovakia	Albania	Algeria	Afghanistan	Philippines	Benin	Burundi	Angola	Argentina	Antigua and Barbuda
Austria	Slovenia	Armenia	Bahrain	Australia	Republic of Korea	Burkina Faso	Djibouti	Botswana	Bolivia	Bahamas
Belgium	Spain	Azerbaijan	Egypt	Bangladesh	Samoa	Cabo Verde	Eritrea	Comoros (Covered by Mauritius)	Brazil (covered by Argentina)	Barbados
Bulgaria	Sweden	Belarus	Iraq	Bhutan	Singapore	Cameroun	Ethiopia	Democratic Republic of the Congo	Chile	Belize
Croatia	Switzerland	Bosnia and Herzegovina	Jordan	Brunei Darussalam	Solomon Islands	Central African Republic	Kenya	Lesotho (covered by SA)	Colombia	Canada
Cyprus	United Kingdom	Georgia	Kuwait	Cambodia	Sri Lanka	Chad	Rwanda	Madagascar	Ecuador	Costa Rica
Czech Republic	San Marino	Israel	Lebanon	China Hong Kong	Taiwan, Republic of China	Congo	Somalia	Malawi	Île de la Réunion	Cuba
Denmark		Kazakhstan	Libya	China Beijing	Thailand	Côte d'Ivoire (Ivory Coast)	South Sudan	Mauritius	Paraguay	Curacao
Estonia		Kosovo	Morocco	Democratic People's Republic of Korea	Timor-Leste	Equatorial Guinea	Uganda	Mozambique	Peru	Dominica
Finland		Krygyzstan	Sudan	Fiji	Tonga	Gabon	United Republic of Tanzania	Namibia	Uruguay	Dominican Republic
France		The FYROM	Occupied Palestine Territory	Heard Island and McDonald Island	Tuvalu	Gambia		Seychelles (Covered by Mauritius)	Venezuela (Bolivarian Republic of)	El Salvador
Germany		Montenegro	Oman	India	Vanuatu	Ghana		South Africa		Grenada
Greece		Republic of Moldova	Qatar	Indonesia	Viet Nam	Guinea		St. Helena		Guatemala
Holy See		Russian Federation	Saudi Arabia	Iran (Islamic Republic of)		Guinea-Bissau (Covered by Cape Verde)		Swaziland (covered by SA)		Guyana
Hungary		Serbia	Syrian Arab Republic	Japan		Liberia		Zambia		Haiti
Iceland		Tajikistan	Tunisia	Kiribati		Mali		Zimbabwe		Honduras
Ireland		Turkey	Yemen	Lao People's Democratic Republic		Mauritania				Jamaica
Italy		Turkmenistan	United Arab Emirates	Macau, SAR of China		Niger				Mexico
Latvia		Ukraine		Malaysia		Nigeria (Lagos + Abuja)				Nicaragua
Liechtenstein		Uzbekistan		Maldives		Sao Tome and Principe				Netherlands Antilles
Lithuania				Marshal Islands		Senegal				Panama
Luxembourg				Micronesia		Sierra Leone				Puerto Rico
Malta				Mongolia		Togo				Saint Pierre and Miquelon
Monaco				Myanmar						St Kitts and Nevis
Netherlands				Nauru						St Lucia
Norway				Nepal						St Vincent and the Grenadines
Poland				New Zealand						Suriname
Portugal				Pakistan						Trinidad and Tobago
Romania				Palau						United States of America
San Marino				Papua New Guinea						United States Minor Outlying Islands