Analysing Human Trafficking
A Machine Learning Approach

2018

INTRODUCTION

Human trafficking is a pervasive and destructive phenomenon in our global society, often underestimated due to its clandestine nature. It may include different acts, for instance the recruitment, transportation or receipt of persons (IOM 2017a). Regardless the act, this human rights violation always has exploitation as its purpose. Exploitation takes various forms like sexual exploitation, forced labour, slavery and others (UNODC, 2008). In many cases, human trafficking involves the crossing of international borders. However, 42% percent of the victims are trafficked within their country of origin (UNODC 2016, p. 9).

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 USD 150.2 billion a year. A variety of measures against the crime exist. Counter-trafficking efforts may include the facilitation of safe and regular migration channels or target the demand side of human trafficking. 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). Finally, to timely identify victims and potential victims of human trafficking, a more comprehensive and clearer understanding of their characteristics is crucial (Aronowitz, 2009; Galos et al., 2017).

Current literature on prevalent victim characteristics mainly draws on categories like age, gender, country of origin and destination (UNODC, 2016; UNODC, 2006). Our aim is to contribute to this strand of research. More precisely, our goal is to detect the most relevant patterns relating the characteristics of victims, their entrance into the trafficking process, and the type of exploitation they are subjected to.

Methodologically, we rely on clustering and classification tree analysis, both machine learning techniques. Our analysis follows a highly inductive approach. That is, we are interested in new insights emerging directly from the data, rather than letting our empirical analysis to be guided by pre-defined theoretical assumptions. Machine learning techniques meet this requirement. Another advantage of machine learning is that it can uncover complex relationships in the data, which is especially valuable for large datasets.

Our main findings corroborate some conventional assumptions and expose yet not considered patterns as to how victim characteristics, their entry into the trafficking process, as well as type of exploitation play together across regions.

The results of the cluster analysis reveal that our data can be grouped into three main victim types. The first subgroup 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.

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, and there is no predefined outcome. In supervised learning, there is an outcome we seek to predict, given the input information we provide. The goal is to formulate a function that best predicts a pre-selected outcome.

The findings of the classification tree analyses illustrate that there are regional differences regarding the prevalent type of exploitation, who is subjected to what kind of exploitation, and how they enter the process of trafficking. In South-Eastern Europe and Central Asia, gender and marital status are defining factors in determining the type of exploitation. For citizens of Asia and the Pacific, the predominant form of exploitation is forced labour, regardless of gender. However, adult women who are recruited through personal contact and exploited in the EU and EEA, West and Central Africa or Southern Africa tend to be subjected to sexual exploitation. For citizens of West and Central Africa, type of exploitation seems to depend on the region of exploitation. While identified victims from this region generally suffer labour exploitation, victims who are trafficked to a continent other than Africa are predicted to be sexually exploited.

The following section presents an overview of the data. Afterwards, the results of the cluster and classification tree analyses are provided. Finally, the last section provides the conclusion.

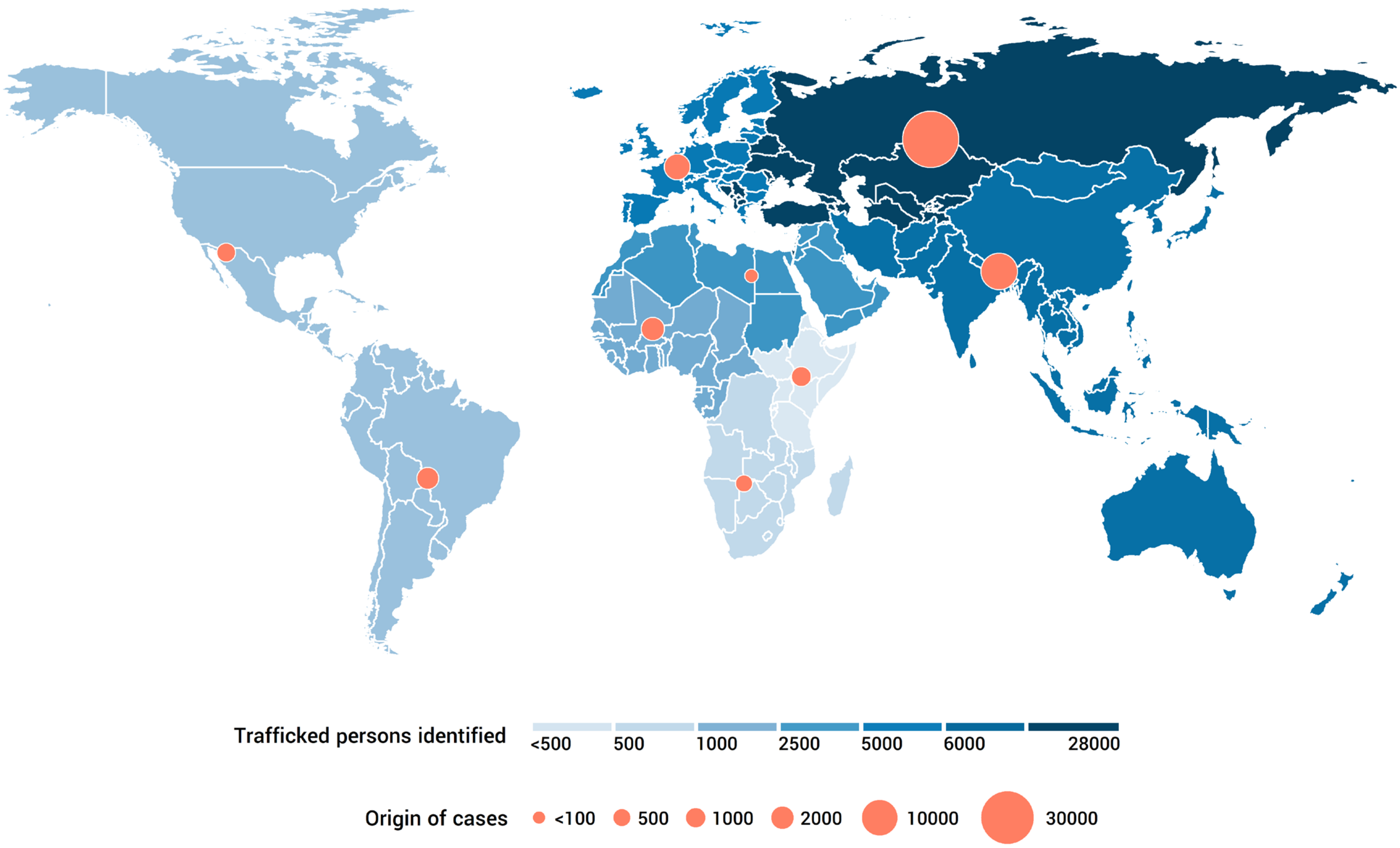

Figure 1. Identified victims of human trafficking by region of citizenship and region of exploitation. Regions are defined according to IOM regional offices.

THE DATA

The individual-level data provided by the International Organization for Migration encompasses over 52,000 identified victims of human trafficking from 2002 to 2017 and is part of IOM's Global Database on Victims of Human Trafficking.

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. We see that human trafficking affects all parts of the world. However, significant differences across regions exist. The number of victims identified range 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).

Most exploitation occurs within the same region as the region of origin (81% in our data) and involves transnational exploitation (58%). Out of the identified victims who were exploited transregionally, 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. An important fact to keep in mind is that the huge variation in the number of identified victims may be due to differences in data collection across regions. Furthermore, a victims' willingness to account their story might be influenced by different cultural norms, levels of trust and privacy concerns, and the sensitivity of the traumatic experiences (Aronowitz, 2009; Galos et al., 2017).

A GENERAL TYPOLOGY OF VICTIMS OF HUMAN TRAFFICKING

What kind of patterns can be derived from the data? To construct a first, very general typology of the victims of human trafficking, we performed cluster analysis, a popular form of exploratory 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 victim profiles.

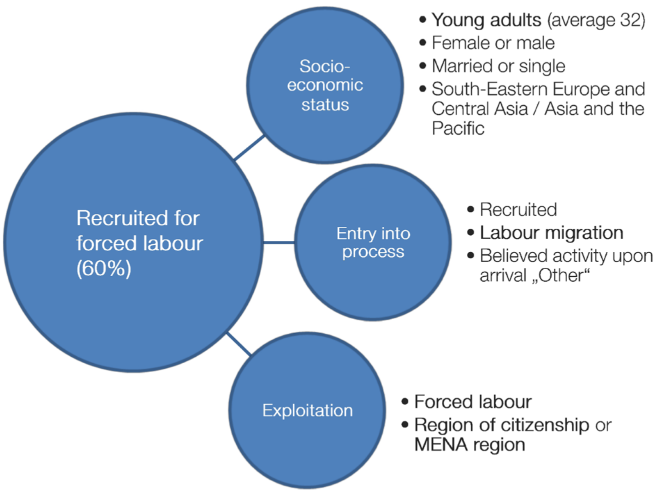

Figure 2. Prototypical characteristics of identified victims of the first subgroup.

GROUP 1: RECRUITED FOR FORCED LABOUR

The first and largest subgroup represents 60% of our 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 the Asia and the Pacific region and are exploited in their region of origin, or in the MENA region (around 15%). Marital status plays a minor role in defining this group, with 44% of the cases being single and 39% married (17% other). The same is true for gender: 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. Literature considers labour migration as a major source of vulnerability, as many of the victims of human trafficking end up in illegal residence situations in their country of exploitation and cannot turn to the authorities for help, due to the illegality of their stay (Clark, 2013). Finally, a striking 71% of 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.

1 Because Cluster Analysis requires a lot of computational power, the analysis was done on a random sample of 1000 observations. 60% of our sample therefore means 600 individuals.

Key Findings Cluster Analysis

- Identified victim of human trafficking in our 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 on human trafficking, as 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.

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 (on average 23) than the victims described in the first group. There is little information regarding 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%). According to the literature, 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 were seeking for work abroad and were deceived and subsequently forced into prostitution (Kligman & Limoncelli, 2005; Surtees, 2008).

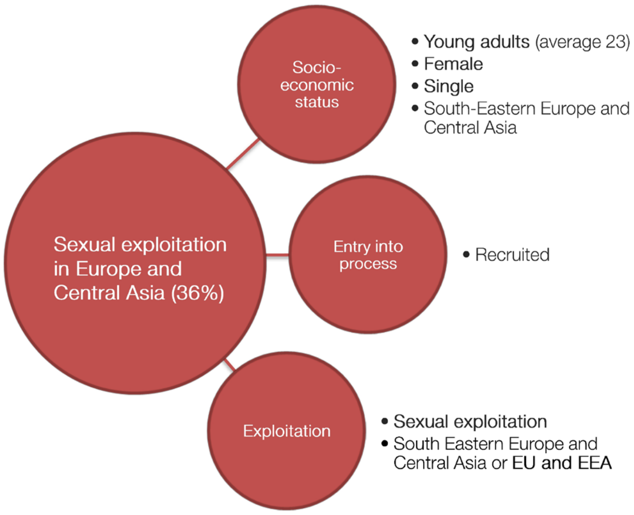

Figure 3. Characteristics of identified victims belonging to Group 2

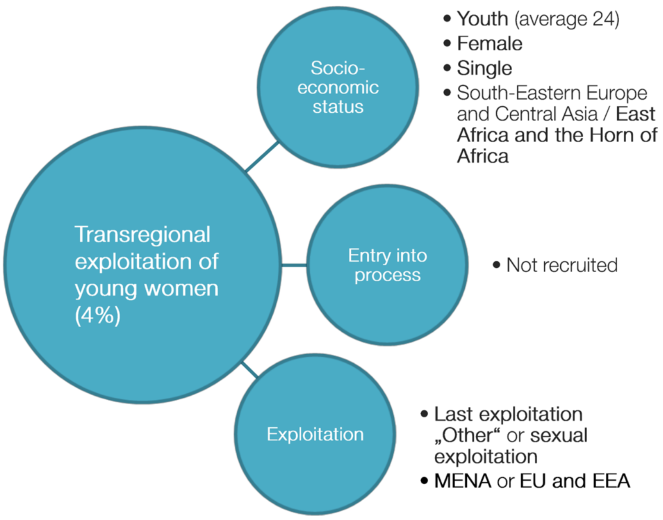

Figure 4. Victim characteristics of the third subgroup

GROUP 3: TRANSREGIONAL 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 transregionally exploited (75%). The victims mainly come 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 was 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 (note that exploitation types are not necessarily mutually exclusive).

Clustering

Cluster analysis is an unsupervised machine learning technique used to detect patterns and groupings in the data. Using a pre-defined 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”.

REGIONAL PATTERNS IN DETERMINING TYPE OF EXPLOITATION

The results from the cluster analysis have shown that dominant dimensions, along which identified victims of trafficking differ, are their origin, their region of exploitation, and the type of exploitation they are subjected to, mostly forced labour or sexual exploitation.

To get a more nuanced picture of the dynamics at work, we took a closer look at how likely a victim, given a set of (pre-exploitation) characteristics, is going to be subjected to either forced labour or sexual exploitation. In other words, we asked ourselves: which victim characteristics best predict the type of exploitation?

To predict type of exploitation, we applied classification tree models, a form of supervised machine learning, useful to uncover complex interactions. Given the provided information on victim characteristics, the model learns how these input characteristics play together in predicting a certain outcome. In our case, 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 we sought to predict was whether an identified victim would be subjected to forced labour or sexual exploitation[1].

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 observations according to type of exploitation, resulting in a "tree". After the tree is constructed, the fitted models were evaluated on the test samples. This means that the rules derived from the training process are tested on data the algorithm has not seen before. Only models that performed well enough (meaning models that predicted the outcomes accurately more than 80% of times) were kept.

Three classifications trees are presented here. The first concerns citizens of South-Eastern Europe and Central Asia. The second tree classifies citizens from Asia and the Pacific. The last tree was estimated on a subset containing citizens from West and Central Africa. 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.

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 (in our case 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. In our case, we relied on the Gini Index to build our 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".

1 In an earlier stage, we also included the outcome category "both labour and sexual exploitation". 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.

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

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

7,696 individuals 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 type of exploitation for citizens of South-Eastern Europe and Central Asia. We see that gender is a very crucial factor. Identified victims who are male or whose gender is not defined have a 98% predicted probability of being a victim of forced labour rather than sexual exploitation. 35% of the observations in our sample follow this pattern, these are around 3,900 cases. Looking at the females, marital status is a defining victim characteristic. For women whose marital status is common law, divorced, separated, or single, and who did not believe that their activity upon arrival was going to be factory work, the predicted probability of being subjected to sexual exploitation is 81%. 48% of our sample (5,404 cases) are classified into this group. By contrast, women who are married or widowed and are recruited 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. We see that 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 our 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

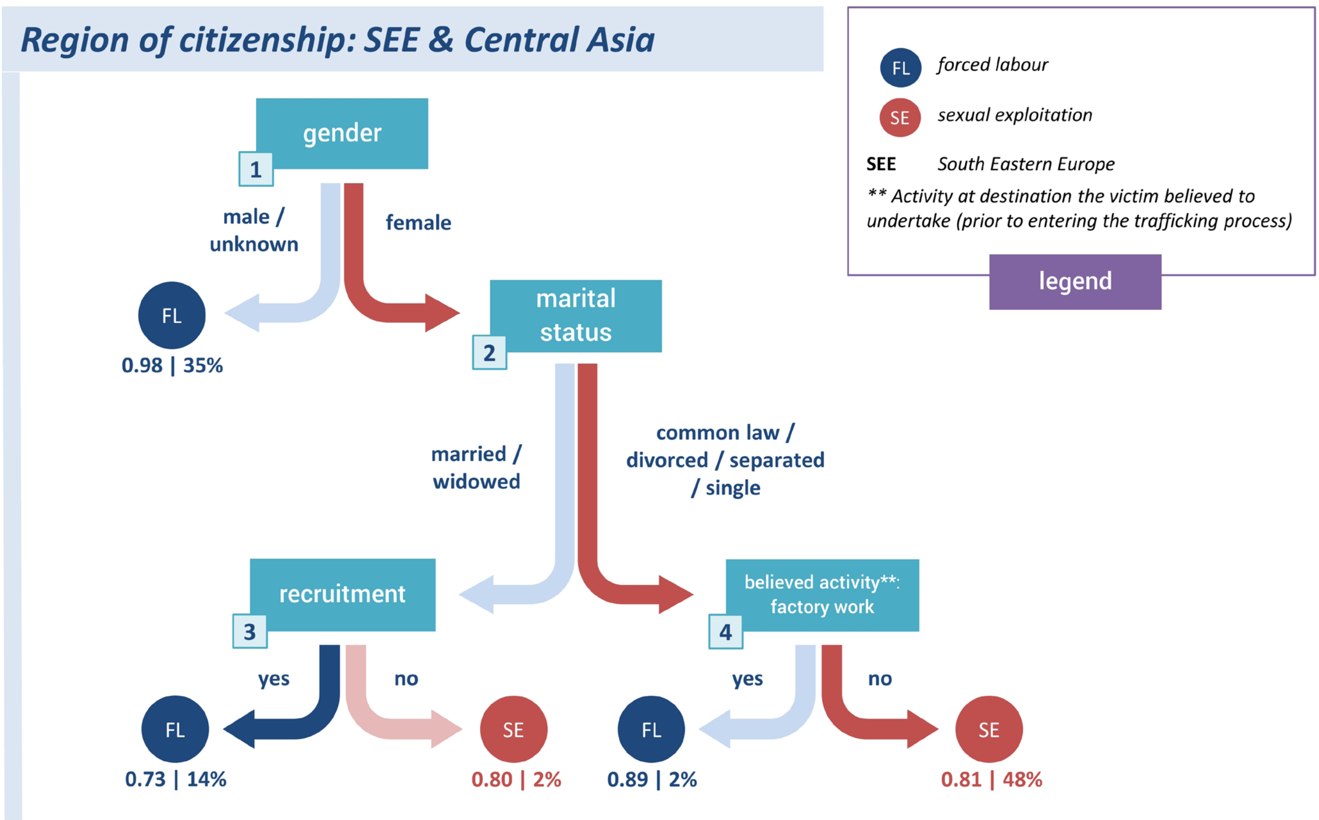

Figure 5. Classification tree predicting type of exploitation for citizens of SEE & CENA

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 gender, 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 include factors like violent marriages or stigmatized status as a widow or single mother (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 reports by IOM caseworkers who conducted interviews with assisted victims of human trafficking.

In general, the female profile depicted in Figure 5 seems to correspond 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 destination country. 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 is recruited or not. 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.

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.

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

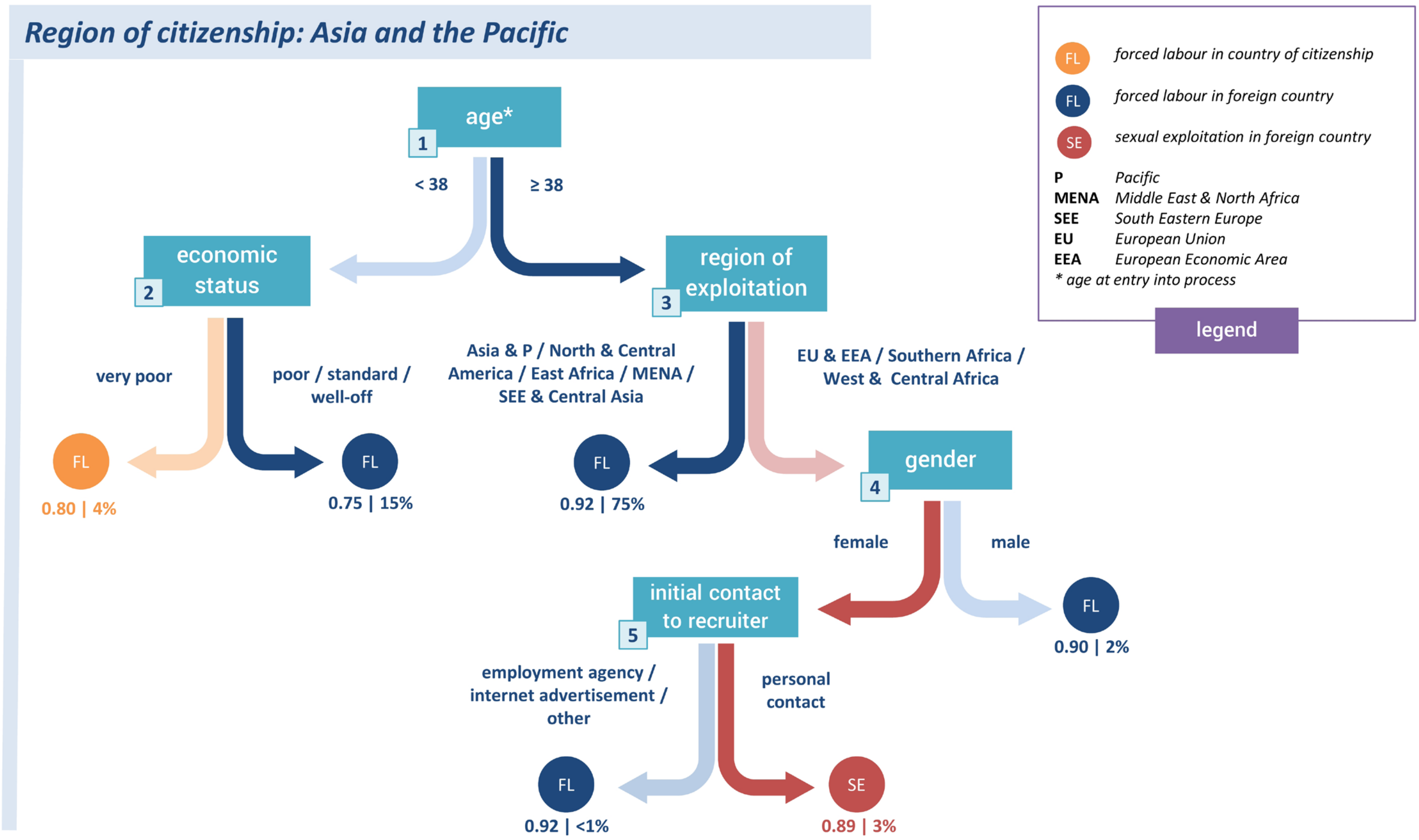

Figure 6. Classification tree predicting type of exploitation for citizens of Asia and the Pacific

Asia and the Pacific constitutes the second largest origin region in our dataset. The most dominant form of exploitation citizens of this region are subjected to is forced labour. Over 5,700 individuals in our 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 type of exploitation is missing.

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 gender is not a defining factor regarding the exploitation type for observations falling into split 1-3. Gender 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 our 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.

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 our 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. Victims who are trafficked within West and Central Africa are predicted (with a 98% predicted probability) to be exploited within the borders of their country of citizenship, if the victim is sold by members of the family. 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 rationalized 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).

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).

"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 them that 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.

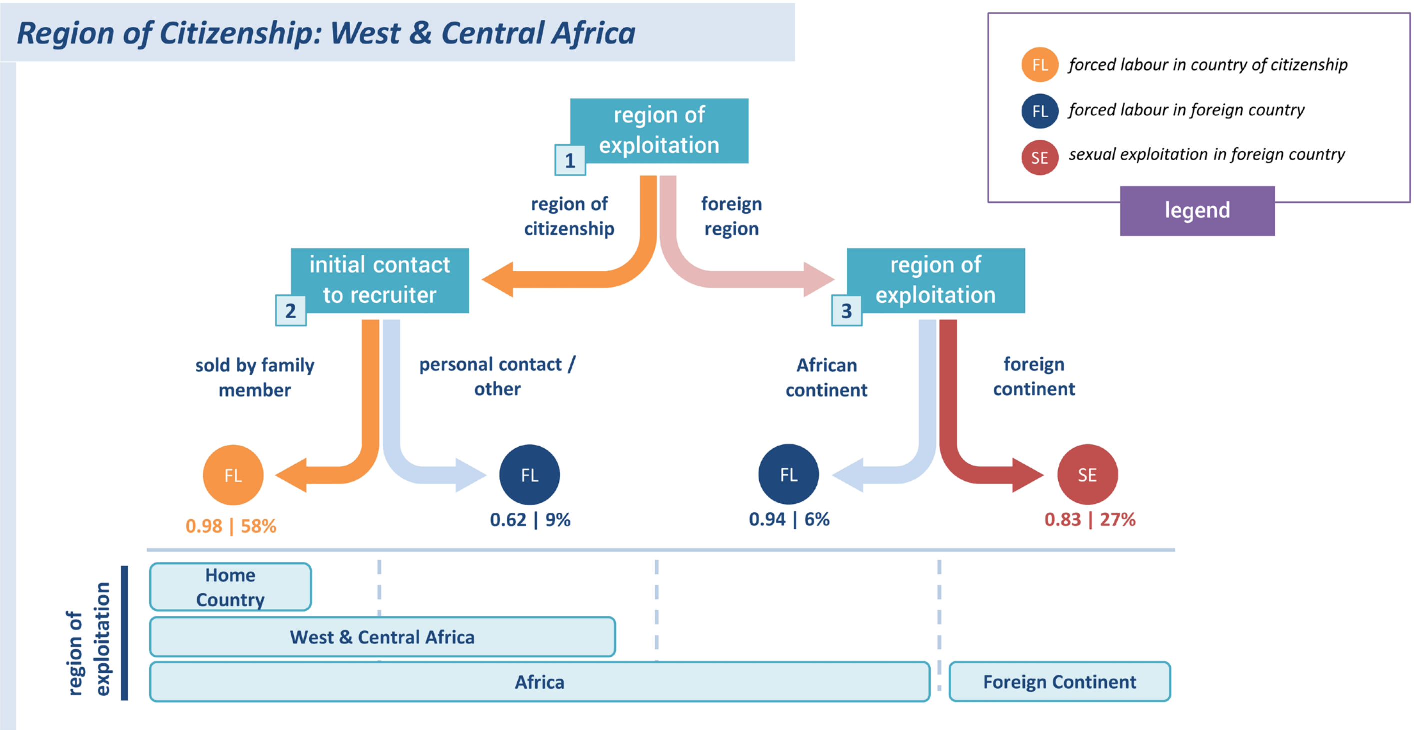

Figure 7. Classification tree predicting type of exploitation for citizens of West and Central Africa

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.

CONCLUSION

Human trafficking affects all regions worldwide and is part of a rapidly growing business of international crime. For prevention and protection in human trafficking, an in-depth understanding of victim characteristics is essential. The goal of our project was to examine more closely which victim characteristics matter the most, depending on how victims enter the trafficking process as well as what kind of exploitation they are subjected to. Aiming to make full use of the individual data provided by IOM, we applied machine learning techniques such as cluster analysis and classification tree models.

In a first step, cluster analysis was applied to generate a very general typology of identified victims of human trafficking. Three prototypical victim types emerged. 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 categorised 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, who are trafficked for exploitation in the MENA region or the EU and EEA. To get a more nuanced picture of the dynamics at work, we conducted classification tree analyses with a focus on type of exploitation. 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, gender proved to be a strong predictor of type of exploitation. Men and victims whose gender is unknown tend to be subjected to forced labour rather than sexual exploitation. For women originating from South-Eastern Europe and Central Asia, marital status can be seen as an indicator predicting type of exploitation. Married or widowed women tend to be recruited for forced labour but there are also married and widowed women who are not recruited and more likely to experience sexual exploitation. Women whose marital status is common law, divorced, separated or single are expected to be trafficked for sexual exploitation but there are also exceptions to this pattern. These are women, who believe that they are going to be working in a factory, who are likely to be exploited for forced labour.

In comparison, citizens from Asia and the Pacific are predominantly exploited for forced labour in a foreign country, regardless of their gender. An exception to this pattern are women who are at least 38 years old, recruited through personal contact and exploited in the EU and EEA, Southern Africa or West and Central Africa. These women tend to be subjected to sexual exploitation. Women who share the same characteristics but are recruited by an employment agency, internet advertisement or other, are more like to be recruited for labour exploitation. Furthermore, victims who are younger than 38 years old and describe their economic status as very poor, rather than poor, standard or well-off, are predicted to be subjected to forced labour in their home country rather than in a foreign country.

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

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

Most importantly, the classification tree models allow 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, our study shows how the application of advanced analysis techniques can be advantageous in the field of human trafficking research. The inductive approach taken enables the identification of victim characteristics moving beyond pre-assumed social groups. Also, through machine learning strategies we were able to map complex interactions of such characteristics. Our approximated models were generated through a learning process directly from the data itself. Classification Tree analysis further has the advantage of dealing well with datasets containing missing values which is of particular interest in the field of human trafficking research.

Consequently, machine learning techniques show 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. We must not forget that our results are based on victims identified and assisted by IOM only, which constitutes a strong selection bias. The precise extent of trafficking remains unknown and there are countless millions of trafficking victims who are never identified (IOM, 2017b).

Keeping these limitations in mind, the results of our analysis are still based on thousands of real victims representing real stories of a considerable amount of men, women, and children around the world. Even though not generalisable, 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, 43*(1-2), 75-98.

Andrees, B., & Van der Linden, M. N. (2005). Designing [1]Trafficking 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-and-analysis/glotip/2016_Global_Report_on_Trafficking_in_Persons.pdf

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

Capstone Couirse, University of Zurich
in collaboration with
International Organization of Migration
2018

Albiez, Janine
Aus der Au, Patrik
Gibilisco, Saro
Woeffray, Théoda
Yin, Jenny