International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Data Analytics on Human Trafficking in India

Rishi Ganji [1], Naga Sucharitha [2], Ashwaq Hussain [3], V. Lingamaiah [4]
[1] B. Tech student, Department of CSE, Anurag Group Of Institutions, Hyderabad
[2] B. Tech student, Department of CSE, Anurag Group Of Institutions, Hyderabad
[3] B. Tech student, Department of CSE, Anurag Group Of Institutions, Hyderabad
[4] Assistant Professor, Department of CSE,Anurag Group Of Institutions, Hyderabad

Abstract-
This paper examines the situation of human trafficking in India. It claims that the discourse on trafficking is still dominated by a focus on either prostitution or illegal immigration, which prioritises state security over human security and fails to address the underlying causes of trafficking as well as the insecurity of those who are victims of it. The root causes or vulnerability factors of trafficking, such as structural inequality, culturally accepted behaviours, poverty or economic insecurity, organ trafficking, bonded labour, and gender violence—all of which are exacerbated by corruption—have not been recognised in academic and policy circles. This essay argues that crimes related to human trafficking that endanger the safety of those trafficked in India require special attention. As a result, it suggests some preventative measures for dealing with and resolving the problem.
Keywords:Trafficking,Crime,Preventative Measures.

I. INTRODUCTION

Human trafficking for sexual exploitation is becoming increasingly common around the world. Trafficking is a significant industry that is growing at the fastest rate in the global illicit economy. This section emphasises the definitions of bonded labour, child labour, and sex trafficking that are used throughout the report in both international and Indian legal standards. Human trafficking for "physical exploitation or any form of sexual exploitation, enslavement or practises equivalent to slavery, servitude, and the forcible removal of organs" is prohibited under the Indian Penal Code's new section 370. Cases involving a wide range of contemporary slavery subtypes have been registered under this provision. Since India joined the Palermo Protocol and changed its penal code, human trafficking for the purpose of sexual exploitation is strictly prohibited. The Immoral Traffic in Persons Act of 1956 also makes it illegal to take, obtain, or influence a person for prostitution. The Protection of Children from Sexual Offenses Act of 2012 makes a number of sexual offences against minors under the age of 18 illegal. Human trafficking has many components, including sex trafficking, labour trafficking, and organ trafficking. Human slavery for prostitution includes sex trafficking as well. Labor trafficking occurs when someone is forced to perform non-sexual labour. A man who has been forced into farm work or a woman who has been forced into domestic servitude are two examples. Finally, organ trafficking is the movement of people in order to sell their organs for transplantation. People may be coerced into trafficking through a variety of means, including physical force or deceptive promises made by traffickers. Promises include false job offers and international marriages. According to the Walk Free Foundation Global Slavery

Index 2014, India is home to an estimated 14 million victims of human trafficking, including victims of sex trafficking, bonded labour, child labour, domestic slavery, and forced marriage, demonstrating that human trafficking is still a global problem. The size of the issue is huge, "both in the number of trafficked victims and increasing number of sites," according to India's 2008 Integrated Plan of Action to Prevent and Combat Human Trafficking. Because of lax law enforcement and limited prosecution, traffickers are motivated by large rewards and low risk. To combat human trafficking, legal action must be taken to recover traffickers' assets and income, as well as to prosecute and punish violators.

II. LITERATURE REVIEW

[1] Comprehensive analysis and meta-[1] Extensive research and meta-analysis. Searches of 15 electronic databases of doctoral theses and peer-reviewed journals now include reference filtering, citation tracking for the included papers, and expert suggestions. Studies were considered if they discussed the possibility of being subjected to violence while being trafficked or the possibility of suffering negative effects on one's physical, mental, or sexual health. Two reviewers independently determined whether papers were eligible and assessed the quality of the research included.

[2] Monitoring, evaluation, and impact assessment principles and techniques have received varying degrees of attention in anti-trafficking activity, but they are frequently the focus of project end evaluations. Evaluations of anti-trafficking programming have primarily focused on assessing project implementation progress and output achievement, rather than tracking achievement of outcomes or impact, which is similar to the findings of reviews of evaluations in the international development sector. The fact that human trafficking is a covert activity, as well as the suffering that its victims endure, complicates matters. As a result, despite some signs of increased awareness and funding, organisations are still struggling to demonstrate their impact and decide which strategies to employ in the fight against human trafficking. This article analyses counter-trafficking programming evaluations generated since the Protocol to draw conclusions about the lessons learned from these interventions and the techniques used to monitor and assess human-trafficking programming.

[3] Since the United Nations declared that increasing awareness should play a significant role in efforts to combat human trafficking, governments and non-governmental organisations have launched a number of public awareness campaigns in an attempt to pique the public's interest and elicit sympathy. These campaigns paint certain people as heroes and villains while downplaying the guilt of others. Governments in destination countries may contribute to the problem of human trafficking by enacting restrictive immigration policies that make migrants vulnerable to traffickers, which is also obscured by this problem portrayal.

[4] The findings of this study suggest that victims' perceptions of specific abilities at different stages of the trafficking experience were influenced by their previous experiences with social support, community support, and their age at the time of crime. Furthermore, as trafficking experience grew, the significance of earlier life events faded. These findings will spark debate among scholars and practitioners about the implications of how power beliefs are

formed, how they evolve over time, and how they may influence service participation following a trafficking experience. With a focus on empowering trafficking survivors and less trauma, the goal of this study was to advance our understanding of more effective methods of prevention, intervention, and after-care services for people affected by this violation of human rights.

[5] The purpose of this paper is to evaluate the outcomes of a Nigerian communication campaign on safe migration in rural areas. The intervention was implemented in 10 secondary schools in Edo State using a means-randomized experiment. The results show that students who received treatment responded appropriately to the information campaign, demonstrating a better understanding of the risks associated with irregular migration and planned behavioural changes. Knowledge of irregular migration grew as a result of receiving precise and comprehensive information about both the risks along the migratory path and the limited reality at the final destination.

[6] The theoretically structured evaluation of fragmented data has additional implications. Many people are initially targeted in campaigns, but only a few receive the message when given the opportunity and incentive to follow the advised behaviour at the end. Furthermore, comments intended to attract attention are frequently misinterpreted, which can have a negative impact on how the general public views both victims and perpetrators. As a result, adopting more narrowly focused behavior-change programmers and designing them with evaluability and evaluation in mind using a learning-oriented approach appears promising. Excellent external evaluations with more resources, as well as low-budget internal evaluations with selective results sharing, could improve our understanding of what truly works and what doesn't.

[7] The goal of this research is to identify significant push and pull variables in human trafficking. They look into the viability of 63 pull factors and 70 push variables that have been published in the literature. They use an extreme bound analysis to accomplish this, running over two million regressions with every conceivable set of variables for as many as 153 countries between 1995 and 2010. According to these findings, crime rates effectively account for both the origin and destination countries of human trafficking. Income has a similar impact, indicating that economic migration and human trafficking have similar root causes. Law enforcement is more important in origin countries than in destination countries. Fascinatingly, gender inequality may impede female mobility, which is critical for the occurrence of human trafficking, and thus may have a restraining effect on human trafficking outflow.

[8] For millions of poor people in South Asia, migration is a critical means of escaping poverty for themselves and their families. They travel long distances in search of work and income to send back home. While many people migrate abroad, others move from rural to urban areas within their own country. This is true for many women and girls who face discrimination at home and want to work in Middle Eastern countries, particularly as domestic workers in private homes. While some migrants are successful, many are vulnerable to being duped by dishonest labour recruiters or exploited by employers in their destination country or region, actions that may constitute forced labour trafficking.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

[9] The Kafala sponsorship system used by Gulf Cooperation Council (GCC) countries employs foreign domestic workers, and this article examines the legal and policy implications of knowledge asymmetry for these workers. Researchers investigate information flows and market uncertainties among five key stakeholders: labor-receiving governments, labor-sending governments, recruitment agencies (subagents), sponsors (employers), and social networks, using ethnographic and field-based observations in major GCC migrant destinations such as Kuwait, Qatar, and the United Arab Emirates (UAE). A lack of bilateral labour agreements, a lack of coordination in government policy, programmes between and among government agencies, a lack of domestic worker labour laws, and the government receiving the labour are all factors that contribute to asymmetric information. For the domestic worker population, these asymmetric information sources create significant market vulnerabilities, which frequently result in job loss and early expulsion. The final section discusses additional policy implications and opportunities for methodological migration research in the GCC.

[10] Complex systems and realism evaluation provide promising methods for evaluating social interventions. Rather than focusing on identifying single causes of observed effects, these techniques consider the intricate interactions between elements to produce results. This study investigates the use of Bayesian networks (BNs) in the realist assessment of interventions to avoid complicated social problems. It employs a theory-based evaluation of the Work in Freedom Programme (WIF), a large anti-trafficking initiative of the International Labour Organization in South Asia supported by the UK.They used BN to study the causal pathways leading to human trafficking using data from 519 Nepalese returnee migrants. According to the findings, the risks of trafficking are primarily influenced by migrants' destination country, recruitment process, and industry in which they work. These findings call into question commonly held beliefs about individual vulnerability and highlight the need for future investments to be made using methods that account for the complexities of an intervention's causal mechanisms in social contexts. BNs are a viable strategy.

II. IMPLEMENTATION

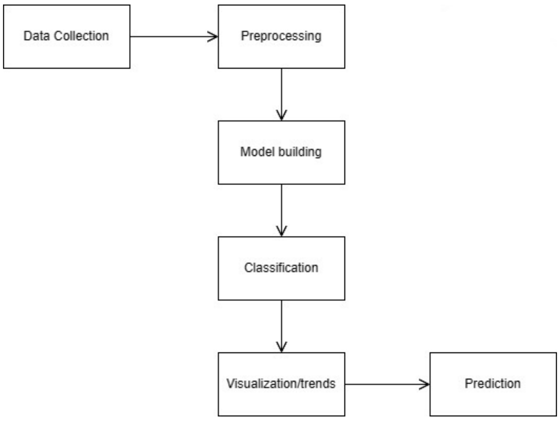


Figure 2.1 System Architecture

Input:-parameters of a person (i.e.:- education, poverty)
Output:- whether he or she will be a victim of human traffic

Module 1: Dataset Gathering

The process of compiling information from various sources in order to advance one's knowledge through "deep learning" is referred to as "data collection." It is critical that the information be saved in a format appropriate for the task at hand.

We will be analysing a CSV dataset on human trafficking in India obtained through Kaggle between 2018 and 2020.

This dataset includes data on the total number of reported cases of human trafficking in each of India's states and

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

union territories, as well as data on the number of victims who were either trafficked or rescued, the nationality of the victims, age groups, the purpose of the trafficking, how the cases were resolved by the police and the courts, and the number of perpetrators who were either arrested or acquitted.

Module 2: Data Preprocessing

Data preprocessing is a data mining technique that converts raw data into a usable and effective format. Data preprocessing converts data into a format that can be processed more quickly and efficiently in data mining, machine learning, and other data science tasks. The techniques are typically applied at the very beginning of the machine learning and AI development pipeline to ensure reliable results. Although there are numerous tools and strategies for data preparation, we have chosen to focus on the following measures:

Finding Null Elements: The dataset we downloaded from Kaggle has some missing elements. This could be because such data was not collected efficiently or because it does not exist. In the dataset, these elements are represented by NaN (not a number) or None. Whatever the reason, it makes our computing more difficult and inaccurate. As a result, we locate missing data and replace it with useful elements.

Pandas treat None and NaN in the same way when indicating missing or null values. To make this convention easier, there are several useful utilities in Pandas DataFrame for finding, removing, and replacing null values. To check for missing values in Pandas DataFrame, we use the functions isnull() and notnull(). Both functions aid in determining whether a value is NaN or not. These functions can also be used with Pandas Series to find null values in a series.

Once we've identified the null values, we can either replace them with mean median mode or remove the rows and columns that contain null values. To remove null values from the dataset, we used the dropna() function. This function removes data columns and rows with null values. When the columns involved have integer or float data types, missing data can be replaced with a mean or median value. The mode value, which is the most frequently occurring value, could be used to replace missing data. This can also be used with floats or integers. However, it is more useful if the relevant columns contain strings. The fillna() function loops through our database, filling in mean, median, and mode values in all empty rows.

Removing Unnecessary Fields: Our dataset contains duplicate information. We believe that removing these features from the dataset is a good idea because their presence has no effect on the target. This method of deleting unnecessary features and keeping only the required features in the dataset is very useful for ensuring that we only work with relevant data.

As a result, we use the drop command to remove a few columns from our dataset. The "df.drop" function in Python can be used to remove a single or multiple columns from a pandas data frame. We also define the column title, index, axis, and labels.

Correlating Features: For a good model, we must correlate each feature with others so that we can determine how our model will behave based on which features. As previously stated, we imported NumPy and specified the two datasets for which we want to check correlation. The Pearson's correlation coefficient is then calculated using the command np.corrcoef

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

(x.y). Finally, we use the print command to print the values.

Module 3: Exploratory Data Analysis

Exploratory Data Analysis, or EDA for short, is a data analysis method that emphasises the use of visual tools. It can be used to identify patterns and trends, as well as validate assumptions, with the help of statistical summaries and graphical representations. The process of data analysis begins with the import of a working dataset or the creation of a new one. Following that is the exploratory analysis phase, which begins immediately. Pandas makes it simple to import a dataset by providing methods dedicated to reading the data. We'll use a variety of Panda capabilities and properties to get a better sense of the big picture. .head() and.tail() are two Pandas functions that are frequently called (). Because of these two, we can display any number of rows, beginning at the beginning or end of the collection.

It is extremely useful for quickly gaining access to a specific section of the dataframe. If we apply the.shape operation to the dataset, Pandas will return a pair of numbers that we can use to represent the dimensionality of the dataset. This feature is quite helpful for gaining a sense of the total number of columns as well as the length of the dataset.

Then, using a variety of Python commands, we create graphs and conduct research on the dataset's category and numerical variables. Following that, we compare and contrast the values of the numeric variables to examine their relationships.

Module 4: Model Training

After our dataset is prepared, we use the following three algorithms to detect human trafficking in the model.

Random Forest

We use it to solve a variety of problems, including regression and classification. It is based on ensemble learning, which combines a number of different classifiers to solve a complex problem and improves the model's performance. A Random Forest is a classifier that takes the average of many decision trees that have been applied to different subsets of a specific dataset in order to improve the accuracy with which that dataset can be predicted. The random forest method does not rely on a single decision tree, but rather aggregates the results of all trees and forecasts the final output based on the forecasts that received the most votes overall.

Decision Tree

A tree can be "learned" by first dividing the source data into subsets and then performing an attribute value test on each subset. This technique is applied to each derived subset in a recursive fashion known as recursive partitioning. The process is complete when the subset at a node all has the same value of the target variable or when splitting no longer adds value to the predictions. This type of knowledge discovery is appropriate for exploratory purposes because it does not require any prior domain knowledge or the specification of any parameters. Decision trees are able to process data with a high dimension. The accuracy of the classifier is generally quite high in decision trees. The induction of knowledge through decision trees is a common type of inductive method used to learn about classification.

Support Vector Machine (SVM)

Support Vector Machine, or SVM for short, is a supervised machine learning technique that can be used for both classification and regression. Although we discuss regression problems, its application is more appropriate for classification. The Support Vector Machine (SVM) algorithm seeks to find a hyperplane in an N-dimensional space that can classify data points separately. The dimension of the hyperplane used is determined by the number of features. If there are only two input features, the hyperplane is simply a straight line. The hyperplane transforms into a two-dimensional plane when three features are used as input. The SVM kernel is a function that takes a low-dimensional input space and converts it to a higher-dimensional input space; in other words, it converts a non-separable problem into a separable problem. The majority of its applications are in the field of non-linear separation. To put it another way, the kernel is in charge of performing extremely complex data transformations before deciding how to partition the data based on the labels or outputs that have been defined.

Training The Models

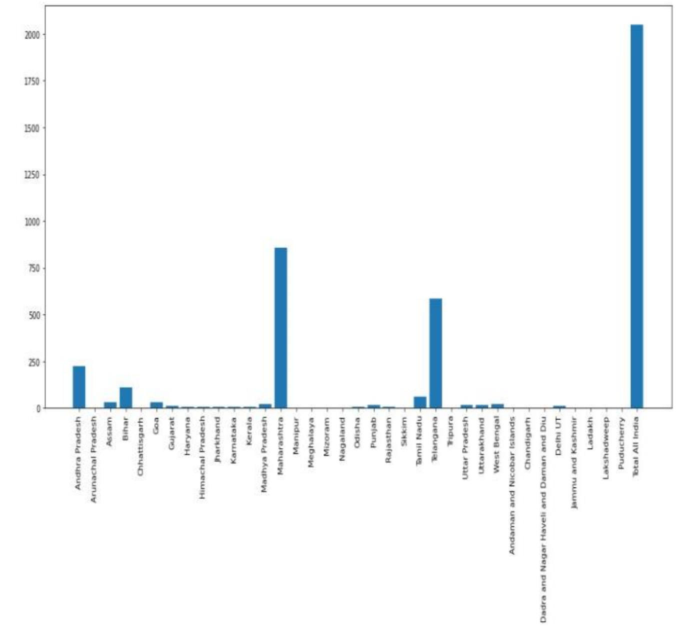
We've already split the dataset into two parts: training data and testing data. We run and validate our training dataset on a CUDA GPU. NVIDIA's CUDA parallel computing framework allows software to use both the CPU and the GPU at the same time. NVIDIA has surpassed AMD to become the most popular GPU provider for cloud computing and machine learning. Furthermore, NVIDIA GPUs are compatible with the vast majority of GPU-enabled Python languages. After loading the data onto the CUDA GPU, we train our model. Because our system has CUDA, we want to move our data from the CPU to the GPU RAM during the training period. As a result, enabling pin memory=True causes the data to be transferred into page-locked memory, which speeds up training. That's all there is to it; the data are now ready for training. After transferring the data to the notebook, a batch size is determined. A data loader is created with the batch size selected.

We evaluate the model and look for changes after it has been trained on a training dataset using CUDA GPU. If there are any changes, we adjust our model and finally confirm the results on test. Finally, we save and run the best model.

IV. RESULT

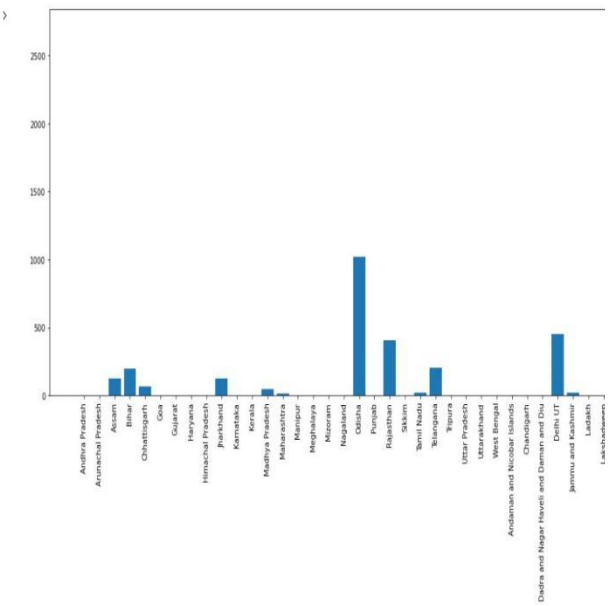
To track human trafficking, the proposed solution employs machine learning algorithms such as SVM, Random Forest, and Decision Tree. As a result, the approach is ideal for text classification problems where access to a dataset with a few thousand tags on each sample is limited. Because the Random Forest emerges as the best model with the highest accuracy, it is used to predict the results.

Because of random forest's sensitivity to unbalanced data, the model's bias was immediately apparent in the initial results. We created a sampling factor to address this issue and provide more balanced data to the model. The proposed algorithm, on the other hand, employs more reasonable logic that is motivated by real-world data.

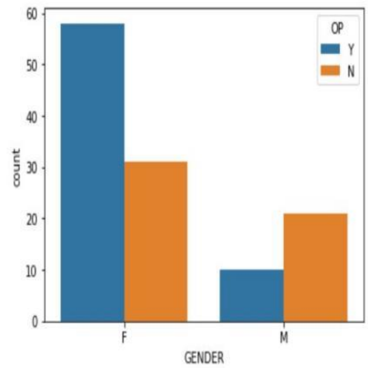


From the Above Graph we can see that Maharashtra and Telengana are top 2 states and Union territories like Andaman and Daman Diu has almost no crime rate in this segment

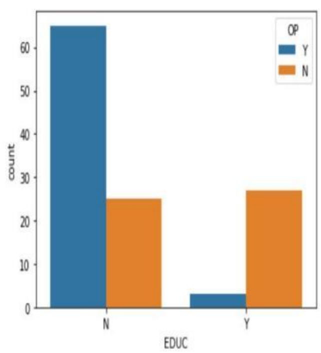
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



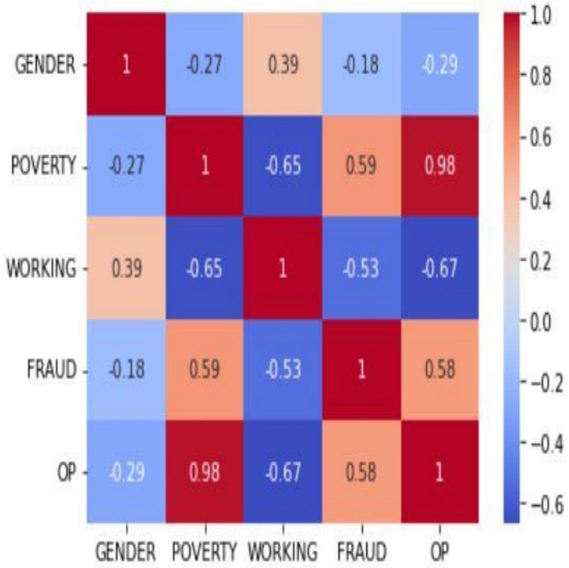
As we can see Odisha Rajasthan is having Higher Rate of Forced Labour than other provi



We can see That the Mostly Females are victim of Human trafficking rather that Male



It is very clear That the persons who has not education has been a victim of than those who has Education



V. CONCLUSION

Human trafficking is a problem that affects all nations, societies, and local communities. This is a widespread issue, especially during a pandemic. Despite the fact that analytics can make a significant contribution, the study of human trafficking in the field of data science is still in its infancy. This paper evaluated previous research on analytics in human trafficking. A structured process was used to provide a clear overview of the findings. Surprisingly, a review of relevant works revealed that a few topics remain unaddressed, particularly the issue of forced labour. Machine learning was not used to detect or forecast forced labour. Although it has been stated that the amount of data available in the field of human trafficking is limited, making analytics difficult to apply, ingenuity can be seen in previous works that have taken advantage of whatever open-source data has been made available to address this issue. As a result, future research should carefully consider how to use any available data to develop an original strategy.

VI. REFERENCES

[1] L. Ottisova, S. Hemmings, L. Howard, C. Zimmerman, and S. Oram. An updated systematic review of the prevalence and risk of violence, as well as the mental, physical, and sexual health problems associated with human trafficking. doi:10.1017/S2045796016000135. Epidemiol Psychiatr Sci. (2016) 25:317-41.

[2] Bryant, K., and T. Landman. Combating human trafficking since Palermo: what do we know about what works? (2020) 6:119-40. doi: 10.1080/23322705.2020.1690097
Human Trafficking in South Asia: Assessing the Effectiveness of Interventions: Rapid Evidence Assessment, Jensen C, Pocock NS, Oosterhoff P. (2020).

[3] E. O'Brien. Human trafficking heroes and villains: portraying the issue in anti-trafficking awareness campaigns. doi: 10.1177/0964663915593410.
Reducing Vulnerability to Human Trafficking, Archer D, Boittin M, Mo C. An Experiment in Changing Knowledge In Nepal, anti-trafficking campaigns, attitudes, beliefs, and practises (2018)

Victims' Perceptions of Power and Identity Intersections: A Human Trafficking Experience, Preble KM. (2016). Online at https://rc.library.uta.edu/uta-ir/handle/10106/25784 (accessed December 14, 2020).

[5] C. Obi, F. Bartolini, and M. D'Haese. Evaluating the impact of an information campaign on youth irregular migration intentions. In Edo State, Nigeria, a randomised control experiment was carried out. In: African Association of Agricultural Economists' Sixth International Conference (AAAE). Abuja (2019) (2019)

[6] Vogel D, Cyrus N. How effective are campaigns addressing the demand side of human trafficking? (2017)

[7] Cho SY. Modeling for human trafficking determinants: an empirical analysis. doi: 10.17645/si.v3i1.125.

[8] International Labour Organization. Work in Freedom Reducing Vulnerability to Trafficking of Women and Girls in South Asia and the Middle East. Informed migration, fair recruitment, and decent work are all supported (2016).

[9] Malit FT, Naufal G. Asymmetric information under the kafala sponsorship system: effects on the income and employment status of foreign domestic workers in GCC countries. International Migration (2016) 54:76-90. doi: 10.1111/imig.12269

[10] L. Kiss, D. Fotheringhame, J. Mak, A. McAlpine, and C. Zimmerman. The use of Bayesian networks for realist evaluation of complex interventions: evidence for human trafficking prevention. doi: 10.1007/s42001-020-00067-8. J Comput Soc Sci. (2020) 4:25-48.