



# TRACKING VICTIMS: USING DATA SCIENCE IN HUMAN TRAFFICKING PATTERNS

Presented by Ana Lucía García Cantú

# HUMAN TRAFFICKING ELEMENTS



## MOVEMENT

Recruitment, transportation, transfer, harboring or receipt of persons



## MEANS OF CONTROL

Threat or use of force or coercion, abduction, fraud, deception, abuse of power or a position of vulnerability



## EXPLOITATION

Sexual exploitation, forced labor, slavery or similar practices, servitude, organ removal, forced marriage



# HYPOTHESIS

## COUNTRY OF ORIGIN

"One vs other"

## TYPE OF EXPLOITATION

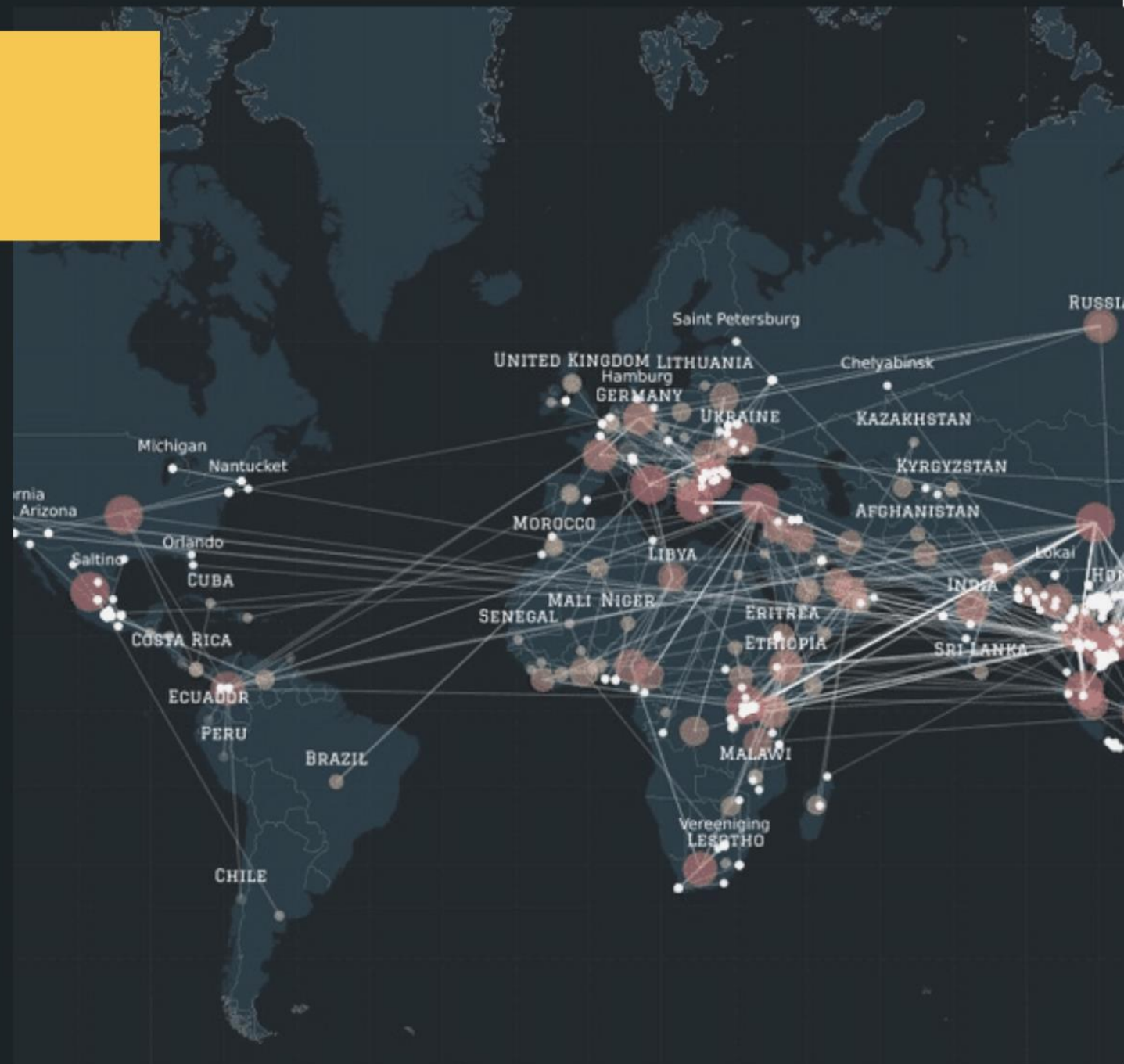
Sexual exploitation

Forced labor

Slavery and similar practices

Forced marriage

## COUNTRY OF EXPLOITATION



# EXPLORATORY DATA ANALYSIS

<https://www.ctdatacollaborative.org/>

| By using this data you agree to the Terms of Use: <a href="https://www.ctdatacollaborative.org/terms-use">https://www.ctdatacollaborative.org/terms-use</a> |                    |            |                 |          |                |                         |               |     |                                |                       |             |                       |                       |                                  |                         |                         |                        |                          |   |
|---|--------------------|------------|-----------------|----------|----------------|-------------------------|---------------|-----|--------------------------------|-----------------------|-------------|-----------------------|-----------------------|----------------------------------|-------------------------|-------------------------|------------------------|--------------------------|---|
|   | yearOfRegistration | Datasource | gender          | ageBroad | majorityStatus | majorityStatusAtExploit | majorityEntry | ... | typeOfSexPrivateSexualServices | typeOfSexConcatenated | isAbduction | RecruiterRelationship | CountryOfExploitation | recruiterRelationIntimatePartner | recruiterRelationFriend | recruiterRelationFamily | recruiterRelationOther | recruiterRelationUnknown |   |
| 0   | NaN                | 2002       | Case Management | -99      | -99            | -99                     | -99           | -99 | ...                            | -99                   | -99         | -99                   | Unknown               | -99                              | 0                       | 0                       | 0                      | 0                        | 1 |
| 1   | NaN                | 2002       | Case Management | -99      | -99            | -99                     | -99           | -99 | ...                            | -99                   | -99         | -99                   | Unknown               | -99                              | 0                       | 0                       | 0                      | 0                        | 1 |
| 2   | NaN                | 2002       | Case Management | -99      | -99            | -99                     | -99           | -99 | ...                            | -99                   | -99         | -99                   | Unknown               | -99                              | 0                       | 0                       | 0                      | 0                        | 1 |
| 3   | NaN                | 2002       | Case Management | -99      | -99            | -99                     | -99           | -99 | ...                            | -99                   | -99         | -99                   | Unknown               | -99                              | 0                       | 0                       | 0                      | 0                        | 1 |
| 4   | NaN                | 2002       | Case Management | -99      | -99            | -99                     | -99           | -99 | ...                            | -99                   | -99         | -99                   | Unknown               | -99                              | 0                       | 0                       | 0                      | 0                        | 1 |

5 rows × 63 columns

1

## REDUCED

Reduced features from 63 to 19

2

## SUBSTITUTED

Homogenized all null values

3

## DIVIDED

Type of exploitation:

- "Forced labour;Sexual exploitation;Combined sexual and labour exploitation"
- "Forced labour;Slavery and similar practices"

4

## TRANSFORMED

From categorical to binary

# COUNTRIES OF ORIGIN

## TOP COUNTRIES

Philippines 11433

Ukraine 7793

Moldova 7499

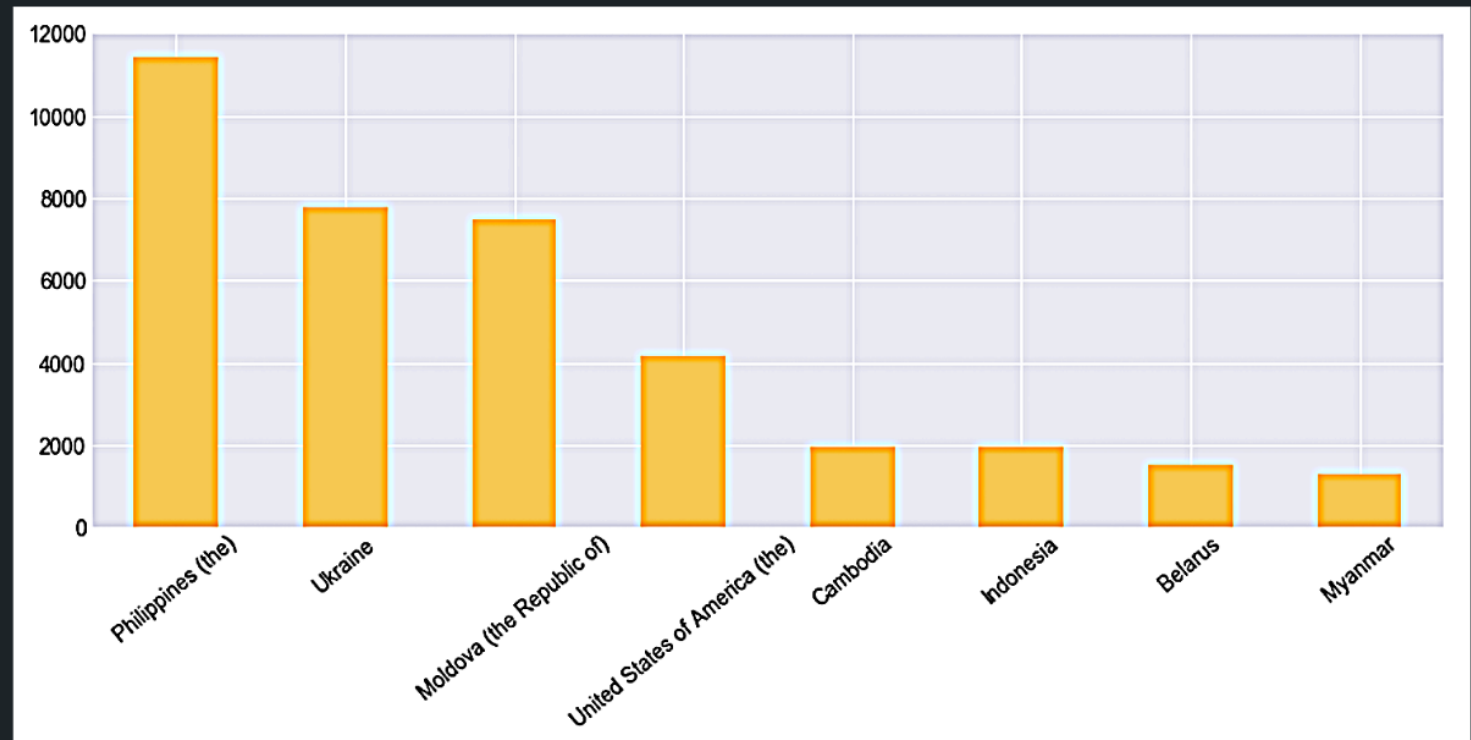
USA 4159

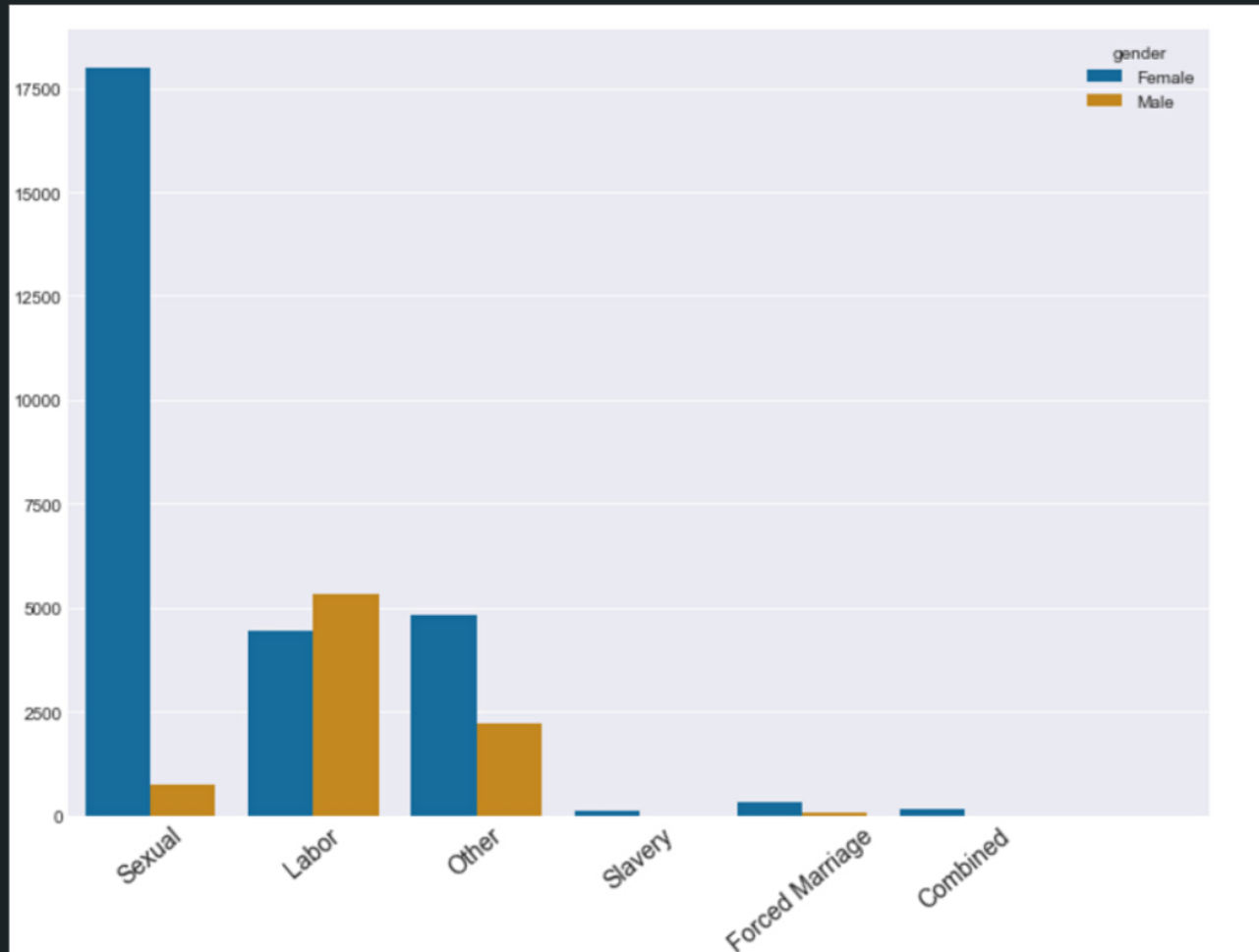
Cambodia 1949

Indonesia 1947

Belarus 1517

Myanmar 1281



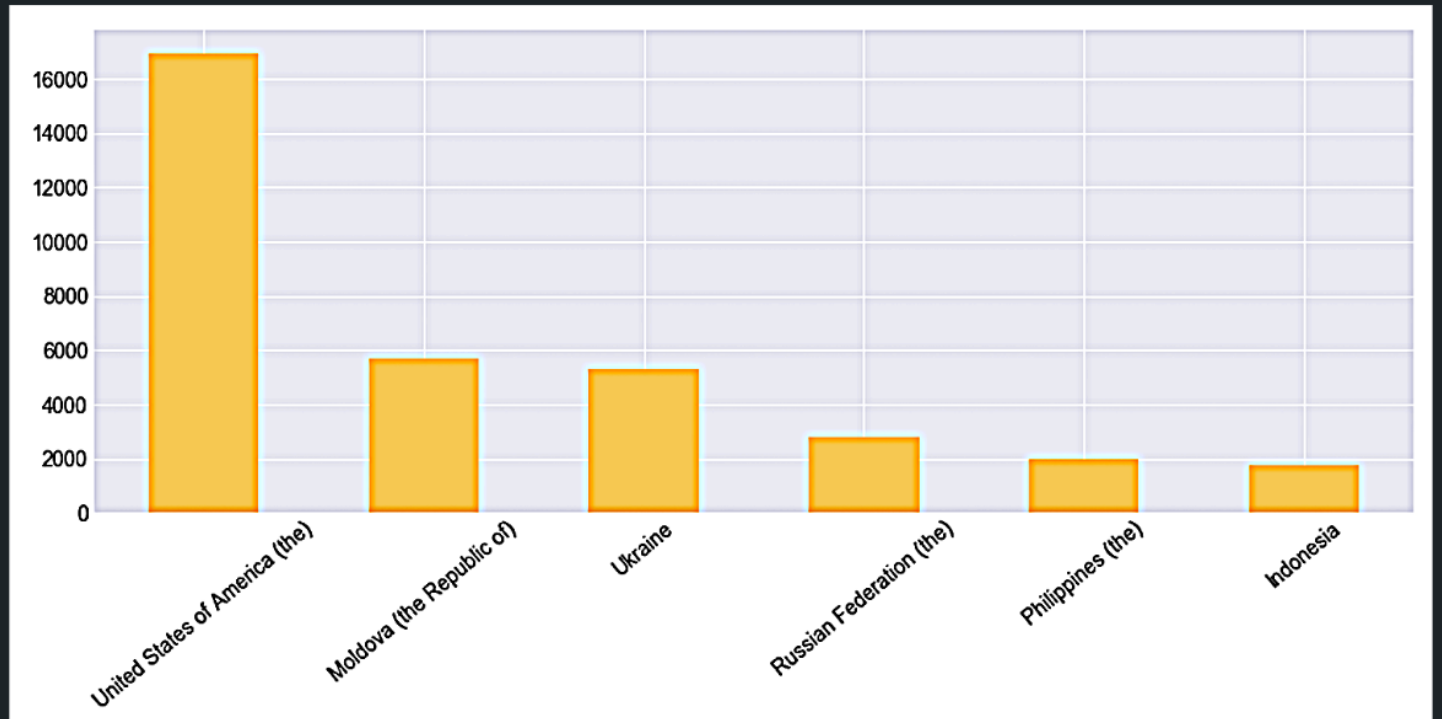


# TYPES OF EXPLOITATION

# COUNTRIES OF EXPLOITATION

## TOP COUNTRIES

USA 16956  
Moldova 5698  
Ukraine 5308  
Russia 2787  
Philippines 1988  
Indonesia 1776



# HYPOTHESIS TESTING

## PREDICTION FOCUS

### KRUSKAL WALLIS TEST

- Three types of classes: country of origin, type of exploitation, country of exploitation.
- Independent variables
- Non-parametric distribution

### STAGE 1

From 58 countries registered, the focus will be on the two top countries, the US and Moldova, and the rest will be encapsuled in "other".

### STAGE 2

Developing a comparison between one country vs "other".



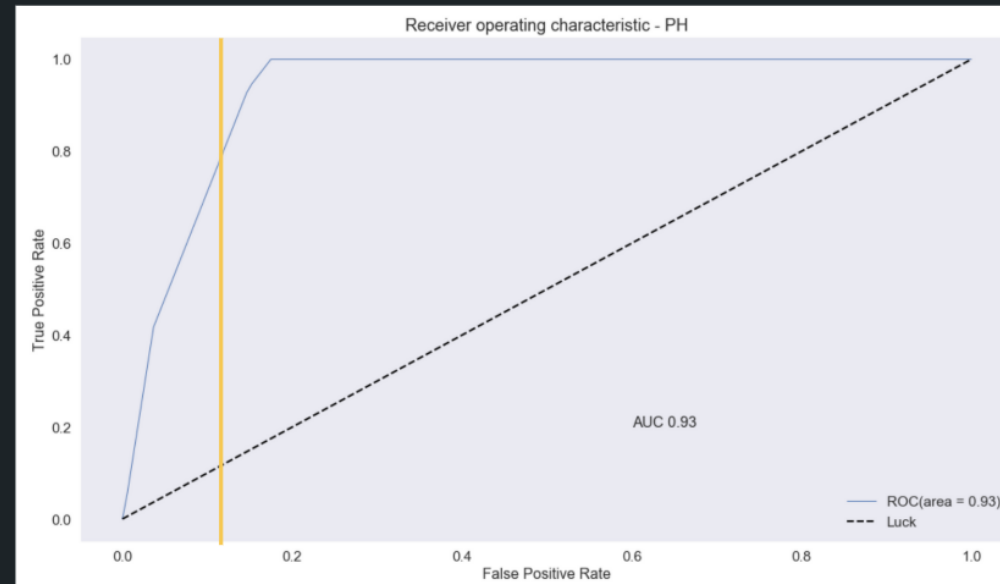
# RANDOM FOREST CLASSIFIER

## THE PHILIPPINES

BASELINE: 0.85648315174255

ACCURACY: 0.9272030651340997

|   | fpr      | tpr      | threshold |
|---|----------|----------|-----------|
| 0 | 0.000000 | 0.000000 | 1.333561  |
| 1 | 0.005698 | 0.055556 | 0.333561  |
| 2 | 0.036665 | 0.417625 | 0.306183  |
| 3 | 0.146285 | 0.927203 | 0.139500  |
| 4 | 0.152358 | 0.946360 | 0.128269  |
| 5 | 0.174852 | 1.000000 | 0.072925  |
| 6 | 1.000000 | 1.000000 | 0.000000  |



|       | predict  | true | predict_binary |
|-------|----------|------|----------------|
| 0     | 0.000094 | 0.0  | 0              |
| 1     | 0.141710 | 0.0  | 1              |
| 2     | 0.000360 | 0.0  | 0              |
| 3     | 0.000323 | 0.0  | 0              |
| 4     | 0.000195 | 0.0  | 0              |
| ...   | ...      | ...  | ...            |
| 13854 | 0.141710 | 1.0  | 1              |
| 13855 | 0.000728 | 0.0  | 0              |
| 13856 | 0.000673 | 0.0  | 0              |
| 13857 | 0.000864 | 0.0  | 0              |
| 13858 | 0.000701 | 0.0  | 0              |

## ROC CURVE

Threshold: 0.139500

## PREDICTIONS

# LOGISTIC REGRESSION

## THE PHILIPPINES

BASELINE: 0.8558337542391226

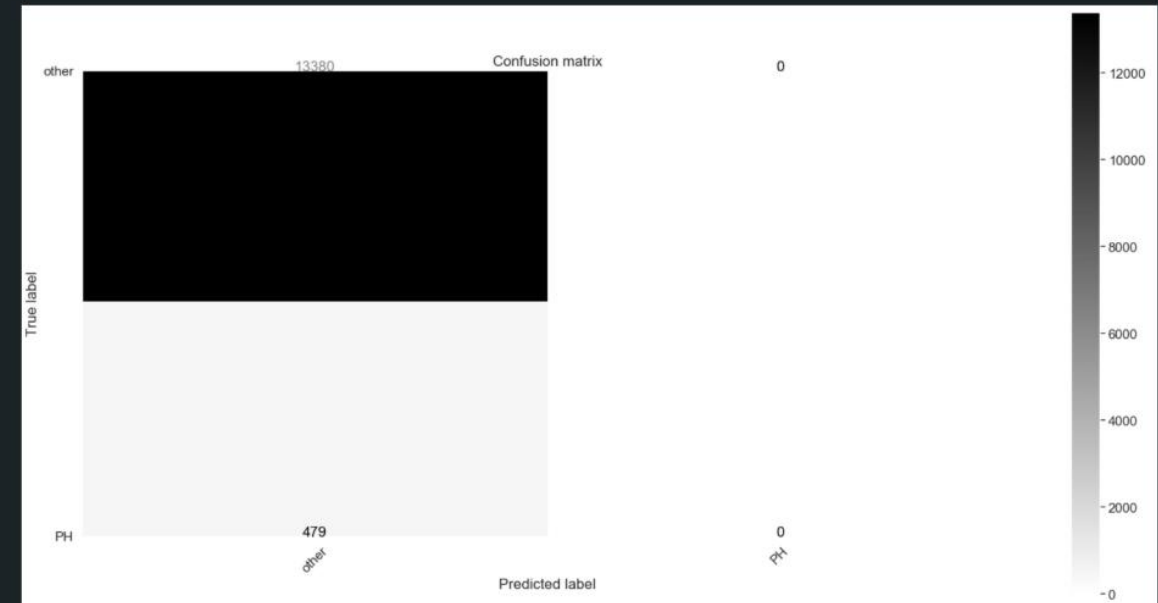
ACCURACY: 0.9337349397590361

PRECISION: 0.93

RECALL: 0.96

|   | fpr      | tpr      | threshold |
|---|----------|----------|-----------|
| 0 | 0.000000 | 0.000000 | 1.323731  |
| 1 | 0.005698 | 0.055556 | 0.323731  |
| 2 | 0.036665 | 0.417625 | 0.302168  |
| 3 | 0.042738 | 0.436782 | 0.147728  |
| 4 | 0.152358 | 0.946360 | 0.136791  |
| 5 | 0.174852 | 1.000000 | 0.073028  |
| 6 | 1.000000 | 1.000000 | 0.000000  |

|       | predict  | true | predict_binary |
|-------|----------|------|----------------|
| 0     | 0.000206 | 0.0  | 0              |
| 1     | 0.003860 | 0.0  | 0              |
| 2     | 0.000325 | 0.0  | 0              |
| 3     | 0.000863 | 0.0  | 0              |
| 4     | 0.139250 | 0.0  | 1              |
| ...   | ...      | ...  | ...            |
| 13854 | 0.289220 | 0.0  | 1              |
| 13855 | 0.303144 | 0.0  | 1              |
| 13856 | 0.000325 | 0.0  | 0              |
| 13857 | 0.000116 | 0.0  | 0              |
| 13858 | 0.139250 | 0.0  | 1              |



## PREDICTIONS

Threshold: 0.136791

## CONFUSION MATRIX

True Negatives (Predicted "other", actual "other") = 13380

True Positives (Predicted "PH", actual "PH")= 0

False Negatives (Predicted "other", actual "PH") = 479

False Positives (Predicted "PH", actual "other")= 0

# NEXT STEPS



## REDUCING FEATURES

A different approach could be one where the types of exploitation are considered as determinants to the place of exploitation.



## CHANGING FEATURES

A new correlation could be set between the gender and age of a victim as predictors of the imposed type of exploitation.