*Folia Oeconomica Stetinensia*
Volume 22 (2022) Issue 1
DOI: 10.2478/foli-2022-0001 | 1–17
ISSN (print): 1730-4237 | ISSN (online): 1898-0198
www.wnus.edu.pl/fos

# A BINARY LOGISTIC REGRESSION MODEL FOR SUPPORT DECISION MAKING IN CRIMINAL JUSTICE

Kateryna M. Berezka, Ph.D., Associate Professor
*Department of Applied Mathematics*
*West Ukrainian National University*
*11 Lvivska Str., Ternopil, Ukraine*
*e-mail: km.berezka@gmail.com*
*ORCID: 0000-0002-9632-4004*

Olha Ya. Kovalchuk, Ph.D., Associate Professor
*Department of Applied Mathematics*
*West Ukrainian National University*
*11 Lvivska Str., Ternopil, Ukraine*
*e-mail: olhakov@gmail.com*
*ORCID: 0000-0001-6490-9633*

Serhiy V. Banakh, Doctor of Law, Associate Professor
*Department of Criminal Law and Process*
*Dean of the Faculty of Law*
*West Ukrainian National University*
*11 Lvivska Str., Ternopil, Ukraine*
*e-mail: s.v.banakh@wunu.edu.ua*
*ORCID: 0000-0002-2300-1220*

Stanislav V. Zlyvko, Doctor of Law, Associate Professor
*Department of Administrative, Civil and Commercial Law and Process*
*Academy of the State Penitentiary Service*
*34 Honcha str., Chernihiv, Ukraine*
*e-mail: academy@kvs.gov.ua*
*ORCID: 0000-0003-2732-3144*

Roksolana Hrechaniuk, Doctor of Law, Associate Professor
*Department of Constitutional, Administrative and Financial Law*
*West Ukrainian National University*
*11 Lvivska Str., Ternopil, Ukraine*
*e-mail: roksolana.gr@gmail.com*
*ORCID: 0000-0003-0566-6365*

**Received 26.02.2022, Revised 22.03.2022, Accepted 28.03.2022**

## Abstract

**Research background:** The economics of incarceration is having an increasing impact on the economies of the world due to the rapid growth in the number of prisoners in the world The search for effective solutions that can help reduce government spending on prisoners in penitentiaries and at the same time ensure the safety of society is becoming increasingly important. These studies used the method of binary logistic regression to predict the probability of convicted criminal recidivism in the future.
**Purpose:** The aim of the paper is to build an effective forecasting model that, based on the statistical and dynamic data of convicts, will provide information for optimal post-trial decisions, such as the grounds for possible parole, probation or length of sentence.
**Research methodology:** The data were collected on the basis of statistical data of 13,010 convicts serving sentences in penitentiary institutions in Ukraine. To predict the probability of convicts committing criminal offenses binary logistic regression and ROC-analysis (Receiver Operator Characteristic analysis) were used .
**Results:** A qualitative binary logistic regression model has been constructed, with the help of which it is possible to predict the probability of criminal recidivism by each of the convicts on the basis of its individual values of the variables included in the model.
**Novelty:** For the first time in Ukraine, a model has been developed to predict the probability of convicts committing repeated criminal offenses.

**Keywords:** economics of incarceration, decision-making, criminal justice, logistic regression, ROC-analysis

**JEL classification:** C52, C59, K42

## Introduction

Today, more than 11 million people are imprisoned worldwide, nearly a quarter of whom are serving sentences in prisons in the United States (World Prison Brief, 2021). That is over 2.5 million. 70 percent of former prisoners are re-incarcerated and the number of prisoners is increasing exponentially, with the cost of such mass imprisonment being enormous. A prisoner's upkeep and infrastructure maintenance can cost up to $ 60,000 a year, so taxpayers' annual costs must exceed this amount (Clay, Phillips, 2013). The consequences of investing such large sums of money in the incarceration of prisoners, rather than in public education or health care systems, are dramatic and devastating for the world's economies.

It is always cheaper to control a person in society than to send him/her to prison. Convicts reimburse only a small part of the considerable state expenses for their detention in penitentiary institutions. To reduce the cost of this process, most progressive countries are actively implementing the use of probation. Thus, those who have committed minor crimes can

leave their prisons and start socializing among law-abiding citizens. Even intensive probation models, such as a combination of rehabilitation and education programs with partial house arrest accompanied by monitoring, are generally cheaper than the most spartan conditions of detention. Across Europe, the cost of probation and the cost of imprisonment range from 1 : 10 to 1 : 25. Ukraine is also joining this process. You can save a lot more by constantly reducing the use of prisons, avoiding the construction of additional facilities or even closing existing ones. A number of scholars from around the world are studying various aspects of the economics of incarceration (Chirakijja, 2018; Grawert, Craigie, 2020; Economics of Incarceration, 2021; McLaughlin et al., 2016; Mumford, Whitmore Schanzenbach, Nunn, 2020).

Today, Ukraine needs to create an effective criminal justice system. A system that punishes for committing an illegal act, but at the same time allows for rehabilitation, is not only more progressive, but also desirable for all social sections of society. In addition, passing a sentence that will allow a person to stay in society will mean that he or she will be able to support himself or herself and his or her family through employment, and thus make a direct economic contribution to the development of society. Probation is expected to save significantly, but the safety of citizens is a priority. Society must be protected even if certain categories of criminals are not behind bars. The most serious criminals will, of course, still be imprisoned. Ensuring public safety must be central to any probation. The new probation model in Ukraine will use risk assessment and management tools. This will allow them to detect and control them better.

In a developed smart society in all spheres of human activity, any systemic, balanced, gradual and planned changes are impossible without prior risk assessment and forecasting the future state of the system. The most effective means for this are the tools of mathematical and statistical methods and information technology (Dmytrów, Gdakowicz, Putek-Szeląg, 2019; Stanimir, 2020a; Stanimir, 2020b; Szkutnik, 2020; Bąk, Wawrzyniak, Sobolewski, 2020; Pilch, 2021; Berezka, Berezsky, Masliy, 2013; Kovalchuk, Shynkaryk, Masonkova, 2021). In many developed countries with high crime rates, such as the United States, statistical methods and information technologies such as artificial intelligence and big data are already used to support optimal decision-making in criminal justice. Machine predictions are not always 100% accurate and the cost of error increases sharply when AI is implemented in tasks with too high stakes, such as people's lives and freedom. However, prognostic justice already exists today. AI algorithms are increasingly being used in criminal proceedings to predict possible crimes and assess the likelihood of their recurrence, in courts when posting bail or sentencing, establishing guilt or innocence. The law enforcement agencies of developed countries are increasingly using in their analytical and prognostic activities predictive methods of predicting probable offenders and

victims of crime, predicting the identity of the perpetrators. Using machine learning algorithms to support decision-making in criminal proceedings can be extremely useful for the judiciary as a whole.

One of the most important features of AI is that its algorithms can be effectively used to assess and reduce the number of unjust sentences (Kovalchuk, 2020).

## 1. Economics of incarceration in Ukraine

Today, the total number of convicts and detainees in Ukraine is about 52,000. According to the report on the use of funds from the General Fund for the Execution of Sentences by institutions and bodies of the State Penitentiary Service of Ukraine, the expenditures of this institution in 2019 amounted to $ 822,774.81, which is approximately $ 24,932.57 per convict. Expenditures on wages and salaries amounted to $ 687,437.98, or $ 20,831.45 per convict. Expenditures on items, materials, equipment and inventory, food, payment for services (except utilities), travel expenses amounted to $ 28,169.22, or $ 853.61 per convict. Expenditures for utilities and energy amounted to $ 75,375.97, or $ 2,284.12 per convict. Expenditures on research and development, individual measures for the implementation of state (regional) programs and other current expenditures amounted to $ 11,593.02, or $ 351.30 per convict. The capital expenditures of penitentiary institutions of Ukraine in 2019 amounted to $ 20,198.61, or $ 612.08 per convict. In addition, in 2019, the Institution incurred expenditures for the training of workers in vocational and technical institutions of social adaptation at penitentiary institutions for a total amount of $ 4,748.06. At the same time, the institution's revenues to the special fund amounted to only $ 42,482.93 (5% of all expenditures of penitentiary institutions under the general fund) (Vysotska, Grechanyuk, Nesin, 2022).

One of the primary tasks of the penitentiary system of Ukraine is to reorientate the essence of the activities of the penitentiary service to punish the criminal to protect society by creating conditions for correction and resocialization. Achieving practical results in the reorganization is possible only after the implementation of systematic, balanced, gradual and planned steps. First of all, this requires the use of a scientific approach and the use of mathematical and statistical methods, including the construction of predictive models to predict the likelihood of convicted criminal recidivism in the future.

We used the binary logistic regression method to identify convicts at low risk of recidivism in order to establish the possibility of a conditional sentence or parole. The method is based on predicting the likelihood of a defendant committing a crime in the future and is designed to

inform judges of the recommended length of sentence. We used static risk factors such as age at the first arrest and gender in combination with dynamic risk factors such as marital status, employment status and education level to determine a comparative risk ratio based on previous offender data.

## 2. Research methodology

The analyzed dataset as many others applied research in different fields includes binary signs. Logistic regression analysis is used in studies of variable dependence from one or several independent variables of an arbitrary type. The ROC-analysis is used in the determination of classification problems based on logit regression.

Logistic regression expresses data mining in the form of dependence $P\{Y = 1 \mid X\} = f(X)$, to predict event rate $\{Y = 1\}$, specified by the values of independent variables $X_1$, $X_2$, ..., $X_k$. The task of logistic regression is to get the prognosis of a continuous variable with values on the segment [0, 1] at any values of independent variables instead of the value prediction of the binary variable (Menard, 2002; "Classification: ROC Curve and AUC", 2022).

The model presented by this type of regression is in effect (per se) the distribution of this law; linear combination of independent variables is the function argument.

Ratio of event probability to incredibility, $P/(1 – P)$ is called the odds ratio (OR).

Probit- and logit-models are used in binary value prognosis for qualitative variables assessment, where linear estimation is complicated by a variety of reasons.

Multiple logit or probit-analysis is a natural continuation of binary and arises from the choice of more than two alternatives.

## 3. Data selection and description

The implementation of the practical part of the research was to identify factors that affect the propensity of convicts to re-commit crimes and predict the likelihood of future recidivism. The sample was formed on the basis of statistical data on 13,010 convicts serving sentences in Ukrainian penitentiaries. The following nominal variables were used in the empirical research:

a) Recidivism (the presence of recurrences): 0 – no,1 – yes;
b) Sex: 1 – male, 2 – female;
c) Age: 1 – to 18 years, 2 – from 18 to 30 years; 3 – from 30 to 45 years; 4 – older than 45 years;

d) Age: 1 (age at the time of the first conviction (to the actual degree of punishment)): 1 – up to 18 years, 2 – from 18 to 30 years; 3 – from 30 to 45 years; 4 – older than 45 years;
e) Age: 2 (age at the time of the first conviction (conditional or actual sentence)): 1 – up to 18 years, 2 – from 18 to 30 years; 3 – from 30 to 45 years; 4 – older than 45 years;
f) Marital status: 1 – single, 2 – married;
g) Education: 0 – incomplete secondary, 1 – secondary, 2 – special secondary, 3 – incomplete higher, 4 – higher;
h) Place of residence (place of residence to the actual degree of punishment): 0 – rural area, 1 –urban area;
i) Type of employment (type of employment at the time of conviction (up to actual punishment)): 0 – unemployed, 1 – part-time, 2 –full-time;
j) Early dismissals (availability of early dismissals): 0 – no, 1 – yes;
k) Motivation for dismissal: 0 – no, 1 – yes

and numerical variables:

a) Number of minor children (number of minor children at the time of conviction (to the real extent);
b) Real convictions (number of real convictions);
c) Conditional convictions (number of conditional convictions);
d) Minor crimes (number of minor crimes – crimes for which punishment in the form of imprisonment for a term of not more than two years is provided, or another, more lenient punishment, with the exception of the main punishment in the form of a fine in the amount of more than three thousand non-taxable minimum incomes of citizens);
e) Crimes of medium gravity (number of crimes of medium gravity – crimes punishable by a fine of not more than ten thousand non-taxable minimum incomes or imprisonment for a term not exceeding five years);
f) Serious crimes (number of serious crimes – crimes punishable by a fine of not more than twenty-five thousand non-taxable minimum incomes or imprisonment for a term not exceeding ten years);
g) Particularly serious crimes (number of particularly serious crimes – crimes punishable by a fine of more than twenty-five thousand non-taxable minimum incomes, imprisonment for a term exceeding ten years or life imprisonment).

Recidivism is the dependent variable, other – independent.

## 4. Empirical results and discussion

The practical part of our research is performed in the software SPSS Statistica.

We construct the logistic regression equation for the dependent variable Recidivism using the forward stepwise method. This is a step-by-step method of selecting variables in which the inclusion test is based on the significance of the Lagrange multiplier criterion statistics.

At step zero, the model is not yet built, all "predicted" values of Y are equal to one, so all observations in which Y = 0, "predicted" are true, and observations where Y = 1, – incorrect (Table 1).

Table 1. Beginning block implementation of the binary logistics model
(Constant is included in the model. The cut value is 0.500)

| Observed | | | Predicted | | |
|---|---|---|---|---|---|
| | | | recidivism | | percentage correct |
| | | | 0 | 1 | |
| Step 0 | Recidivism | 0 | 6,508 | 0 | 100.0 |
| | | 1 | 6,396 | 0 | 0.0 |
| | Overall Percentage | | | | 50.4 |

Source: own elaboration using the SPSS Statistica software.

In the next steps, the logistics model will predict the chances and probabilities of unrealized projects based on the independent variables of the regression equation.

The quality of the regression model approximation is estimated using the likelihood function. A measure of plausibility is the negative double value of the logarithm of this function –2log (Table 2). The smaller this value, the better the binary logistics model is built.

Table 2. Model summary for the binary logistics model

| Step | –2 Log likelihood | Cox & Shell R Square | Nagelkerke R Square |
|---|---|---|---|
| 1 | 14,387.754 | 0.238 | 0.317 |
| 2 | 12,266.685 | 0.353 | 0.471 |
| 3 | 11,079.989 | 0.410 | 0.547 |
| 4 | 9,592.810 | 0.474 | 0.632 |
| 5 | 6,683.836 | 0.580 | 0.774 |
| 6 | 2,846.378 | 0.688 | 0.918 |
| 7 | 2,801.256 | 0.689 | 0.919 |
| 8 | 2,793.620 | 0.690 | 0.919 |

Source: own elaboration using the SPSS Statistica software.

The initial value for –2log is the value obtained for a regression model that contains only constants. After adding all significant variables, the value of –2log is 2,793.62 and is much smaller than the initial 14,387.75. Such a reduction in size means an improvement in the model. The difference is denoted as the value of Chi-square is significant (Table 3). This means that the initial model after the addition of influential variables has become much better.

Table 3. Estimates of the binary logistics model

| | | Chi- square | df | Sig. |
|---|---|---|---|---|
| Step 8 | Step | 7.636 | 1 | 0.006 |
| | Block | 15,094.150 | 8 | 0.000 |
| | Model | 15,094.150 | 8 | 0.000 |

Source: own elaboration using the SPSS Statistica software.

**Cox & Snell R Square and Nagelkerke R Square** – pseudo-coefficients of determination, obtained on the the ratio basis of plausibility functions related to models with only a constant and with all coefficients. These coefficients are rarely used when comparing models.

Based on the logistic regression model, it is possible to predict whether or not an event will occur {Y – 1}. The forecast rule, is used by default in the logistic regression procedure: if the probability of an event is greater than 0.5, it is assumed that the event will occur; if less than or equal to 0.5, it is assumed that the event will not occur.

According to the classification table (step is not equal to zero) determine the number of correctly and incorrectly predicted responses in each of the categories of the analyzed variable and the total percentage of correct predictions (Table 4).

Table 4. Binary logistics model classification table
(The cut value is 0,500)

| Observed | | | Predicted | | |
|---|---|---|---|---|---|
| | | | recidivism | | percentage correct |
| | | | 0 | 1 | |
| Step 8 | Recidivism | 0 | 6,402 | 106 | 98.4 |
| | | 1 | 88 | 6,308 | 98.6 |
| | Overall Percentage | | | | 98.5 |

Source: own elaboration using the SPSS Statistica software.

Analysis of the variables of the regression model obtained by the method forward stepwise (Table 5).

Table 5. Model summary for the binary logistic model

| Variables $X$ | | B | S.E. | Wald | df | Sig. | Exp(B) |
|---|---|---|---|---|---|---|---|
| Step 8[h] | Age1 | –0.546 | 0.086 | 40.135 | 1 | 0.000 | 0.579 |
| | Type of employment | –0.206 | 0.075 | 7.592 | 1 | 0.006 | 0.814 |
| | Conditional convictions | –5.449 | 0.114 | 2,273.396 | 1 | 0.000 | 0.004 |
| | Minor crimes | 5.596 | 0.116 | 2,344.087 | 1 | 0.000 | 269.265 |
| | Crimes of medium gravity | 5.750 | 0.116 | 2,462.578 | 1 | 0.000 | 314.216 |
| | Serious crimes | 5.610 | 0.117 | 2,306.734 | 1 | 0.000 | 273.098 |
| | Particularly serious crimes | 5.790 | 0.150 | 1,484.667 | 1 | 0.000 | 327.105 |
| | Early dismissals | 1.687 | 0.134 | 158.962 | 1 | 0.000 | 5.402 |
| | Constant | –8.170 | 0.266 | 942.777 | 1 | 0.000 | 0.000 |

B – coefficients $a_j$ regression equation; S.E. – measure of the variability of the coefficients $a_j$; Wald – criterion of the significance of the regression coefficients. The higher its value (along with the number of degrees of freedom), the higher the significance of the regression coefficient; df – number of degrees of freedom; Sig. – significance according to Wald's criterion (Wald) ($H_0$: $a_j = 0$); Exp(B) – odds ratio $e^{a_j}$ (used to interpret the results).

Source: own elaboration using the SPSS Statistica software.

The observed significance of the coefficients is calculated on the basis of Wald's statistics. Its universality makes it possible to assess the significance of not only individual variables but also categorical variables in general, despite the fact that they are deaggregated into index variables. Wald's statistics have a distribution chi-square. Degrees of freedom is equal to one if the hypothesis of zero coefficient for a regular or index variable is tested, and for a categorical variable is equal to the number of values minus one (the number of corresponding index variables). The square root of Wald's statistics is approximately equal to the ratio of the value of the coefficient to its standard error (as well as *t*-statistics in the usual linear regression model).

In the table of coefficients (Table 5) the variables Age1 (age at the time of the first conviction (to the actual degree of punishment)), Type of employment (type of employment at the time of conviction (up to actual punishment) are significant), Conditional convictions (number of conditional convictions), Minor crimes (number of minor crimes), Crimes of medium gravity (number of crimes of medium gravity) and Early dismissals (availability of early dismissals). Other variables were not included in the regression equation.

The following equation of the regression logistic model is constructed:

$$\hat{Y} = -0.55A - 0.21Te - 5.45Cc + 5.6Mc + 5.75Cm + 5.61Sc + 5.8Ps + 1.69Ed - 8.17$$

where: $A$ – Age1, $T$ – type of employment, $Cc$ – conditional convictions, $Mc$ – Minor crimes, $Cm$ – crimes of medium gravity, $Sc$ – serious crimes, $Ps$ – particularly serious crimes, $Ed$ – early dismissals.

Substituting the corresponding values of the independent variables into this equation will result in a logarithm of the odds ratio (OR) of the recidivism of the offender. To determine the odds themselves, it is necessary to raise the number e (the basis of natural logarithms) to this degree. Calculate the probability that recidivism will be committed in the future, based on information about age at the time of the first conviction (to the actual degree of punishment), type of employment at the time of the conviction (up to actual punishment), number of conditional convictions, number of minor crimes, number of crimes of medium gravity and availability of early dismissals are possible with the help of the following ratio:

$$\mathrm{OR} = \frac{P}{1-P} \tag{1}$$

where $P$ the likelihood of recidivism.

Let us solve this equation relatively $P$:

$$\mathrm{P} = \frac{\mathrm{OR}}{1+\mathrm{OR}} \tag{2}$$

We use equation (2) to predict the chances and probability of committing a criminal offense in the future by a convict who was first sentenced to a real sentence before reaching adulthood (Age1 = 0), unemployed ($Te$ = 0), twice suspended ($Cc$ = 2), committed one minor crime ($Mc$ = 1), one medium crime ($Cm$ = 1), one serious crime ($Sc$ = 1), one particularly serious crime ($Ps$ = 1) in which there were early releases ($Ed$ = 1):

$$Log(OR_{\text{recidivism}}) = \hat{Y} = \tag{3}$$

$$= -0.550 - 0.210 - 5.452 + 5.61 + 5.751 + 5.61 + 5.81 + 1.691 - 8.17 = 5.38$$

$$OR_{\text{recidivism}} = \exp(5.38) = 217 \tag{4}$$

the corresponding probability of recidivism by these convicts $P_{\text{recidivism}} = \frac{217}{1+217} = 0.99$.

Similarly, we calculate the chances and probability of committing a criminal offense in the future by a convict who was first convicted to the actual degree of punishment over the age of 45 (Age1 = 4), worked full time ($Te$ = 2), was not conditionally sentenced ($Cc$ = 0), committed one crime of minor gravity ($Mc$ = 1), one crime of medium gravity ($Cm$ = 1) and did not have early releases ($Ed$ = 0):

$$Log(OR_{\text{recidivism}}) = \hat{Y} = \tag{5}$$
$$= -0.554 - 0.212 - 5.450 + 5.61 + 5.751 + 5.60 + 5.80 + 1.690 - 8.17 = 0.56$$

$$OR_{\text{recidivism}} = \exp(2,25) = 1.75 \tag{6}$$

$$_{\text{recidivism}} = \frac{9.49}{1 \quad 9.49} = 0.64 \tag{7}$$

Thus, the chances of committing a criminal offense in the future by a convict who was first sentenced to a real sentence over the age of 45, had a full-time job, did not commit serious and especially serious crimes and had no early release are reduced from 0.99 to 0.64, which is more 30%. Change $X_j$ per unit causes a change in the odds ratio y $e^{aj}$ times.

If independent variables $X_1, X_2, \cdots, X_k$ are too closely related (the condition of their independence is violated), the constructed regression equation may be incorrect. In our case, there is no dense correlation between the studied variables (Table 6).

Table 6. Correlation matrix of the logistics model

| | Constant | Age1 | Type of employment | Conditional convictions | Minor crimes | Crimes of medium gravity | Serious crimes | Particularly serious crimes | Early releases |
|---|---|---|---|---|---|---|---|---|---|
| Constant | 1.000 | –0.685 | –0.125 | 0.464 | –0.543 | –0.577 | –0.571 | –0.509 | –0.191 |
| Age1 | –0.685 | 1.000 | –0.096 | 0.071 | –0.037 | –0.030 | –0.030 | –0.035 | –0.041 |
| Type of employment | –0.125 | –0.096 | 1.000 | 0.134 | –0.114 | –0.094 | –0.095 | –0.044 | –0.084 |
| Conditional convictions | 0.464 | 0.071 | 0.134 | 1.000 | –0.884 | –0.888 | –0.832 | –0.653 | –0.176 |
| Minor crimes | –0.543 | –0.037 | –0.114 | –0.884 | 1.000 | 0.853 | 0.820 | 0.681 | 0.111 |
| Crimes of medium gravity | –0.577 | –0.030 | –0.094 | –0.888 | 0.853 | 1.000 | 0.839 | 0.705 | 0.136 |
| Serious crimes | –0.571 | –0.030 | –0.095 | –0.832 | 0.820 | 0.839 | 1.000 | 0.689 | 0.142 |
| Particularly serious crimes | –0.509 | –0.035 | –0.044 | –0.653 | 0.681 | 0.705 | 0.689 | 1.000 | 0.125 |
| Early releases | –0.191 | –0.041 | –0.084 | –0.176 | 0.111 | 0.136 | 0.142 | 0.125 | 1.000 |

Source: own elaboration using the SPSS Statistica software.

As a result of the logistic regression, a diagram was formed in which two values were used to indicate the predicted probability: 1 – a recurrence will be committed and 0 – no recurrence will be committed (Figure 1). The values of the predicted probability are plotted on the horizontal axis, and the frequencies are plotted on the vertical axis. There are no columns that contain both 0 and 1, which indicates the absence of errors in the forecast. The numbers 1 in

the right part of the chart and the numbers 0 in the left part correspond to the correct predictions regarding the commission of criminal recidivism.

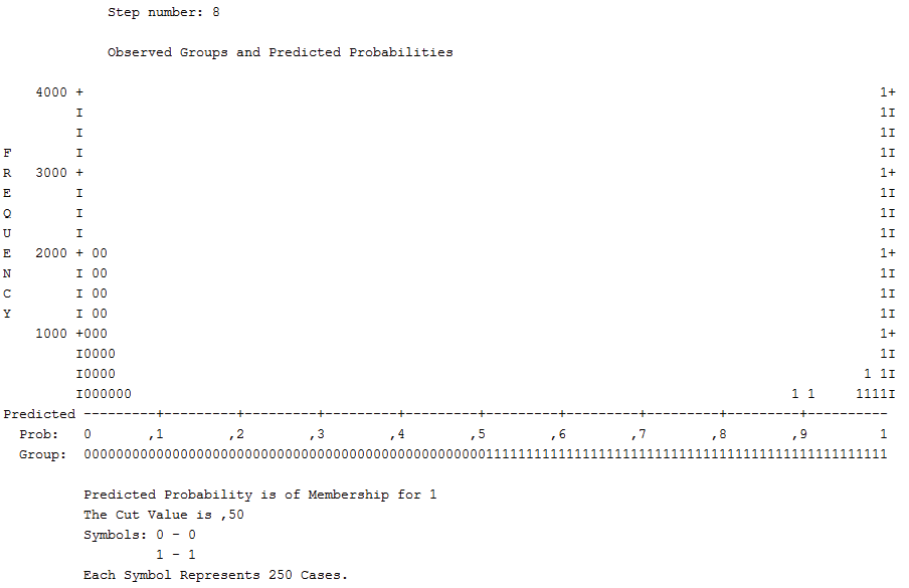



Figure 1. Diagram of the predicted probability of recurrence

Source: own elaboration using the SPSS Statistica software.

To present the results of binary classification and evaluate the effectiveness of classification schemes, we use the procedure of the ROC-analysis ("Classification: ROC Curve and AUC", 2022).

The classification table (Table 4) is based on the results of classification according to the model (cut-off threshold is 0.5) and the actual (objective) affiliation of convicts to different classes – those who had relapses (Positive) or did not commit repeated crimes (Negative).

In the analysis we will not use absolute indicators, but relative – rates, expressed as a percentage.

True Positives Rate

$$TPR = \frac{TP}{TP + FN} \times 100\% \tag{8}$$

False Positives Rate

$$FPR = \frac{FP}{FP + TN} \times 100\% \tag{9}$$

The objective value of each binary classifier is determined by the sensitivity and specificity of the model.

Sensitivity – the proportion of truly positive cases

$$Se = TPR = \frac{TP}{TP + FN} \times 100\% \tag{10}$$

Specificity – the proportion of truly negative cases that were correctly identified by the model

$$Sp = \frac{TN}{FP + TN} \times 100\% \tag{11}$$

$FPR = 100\% - Sp$.

For our research

$$Se = TPR = \frac{TP}{TP + FN} \times 100\% = \frac{6308}{6308 + 88} \times 100\% = 99\% \tag{12}$$

$$Sp = \frac{FP}{FP + TN} \times 100\% = \frac{6402}{6402 + 106} \times 100\% = 98\% \tag{13}$$

The built model has high sensitivity and specificity. It gives a true result if there is a positive result – it detects cases of recidivism that can be committed by convicts and gives a reliable result in predicting the non-recurrence of convicts in the future.

Table 7. Rates for the binary logistic model

| Rate | Rate value in percent |
|---|---|
| True positive | 6308 |
| False positive | 6402 |
| True negative | 106 |
| False negative | 88 |

Source: own elaboration.

The quality of the constructed model was evaluated by estimating the area under the ROC curves. Theoretically, the area varies from 0 to 1, but since the model is always characterized by a curve located above the main diagonal, we usually talk about changes from 0.5 (“useless” classifier) to 1 (“ideal” model).

A graph, also known as an error curve, allows you to estimate the quality of a binary classification. It reflects the ratio between the share of objects from the total number of bearers of the feature, correctly classified to the total number of objects that do not carry the feature, erroneously classified as having a feature. ROC-analysis allows us to conclude that the classification is qualitative: the ROC-curve of the binary classification is located above the main diagonal model (Figure 2).

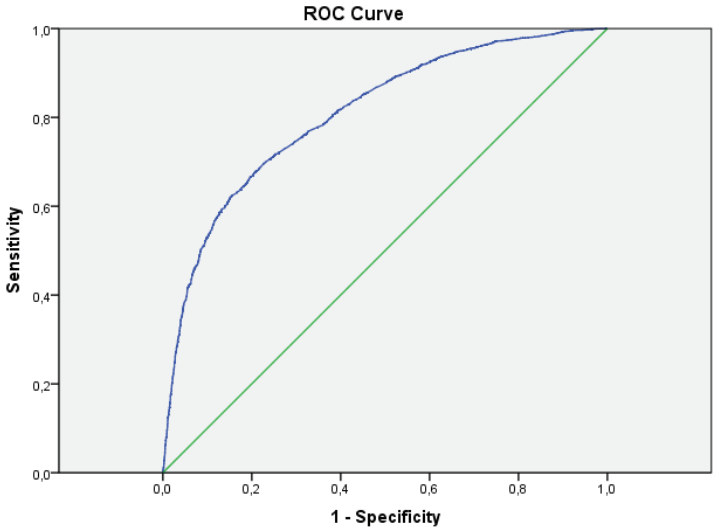



Figure 2. ROC – the curve of the binary classification prediction of recurrence

Source: own elaboration using the SPSS Statistica software.

This estimate can be obtained directly by calculating the area of the polygon under the experimentally obtained ROC curve. The numerical indicator of the area under the curve is called AUC (Area under the curve) (Table 8).

Table 8. Estimation of the area under the ROC curve

| Area | Std. error[a] | Asymptotic Sig.[b] | Asymptotic 95% confidence interval | |
|---|---|---|---|---|
| | | | lower bound | upper bound |
| 0.809 | 0.004 | 0.000 | 0.801 | 0.816 |

[a] Under the nonparametric assumption.
[b] Null hypothesis: true area = 0.5).

Source: own elaboration using the software SPSS Statistica.

It is believed that the higher the AUC, the greater the predictive power of the model. However, the AUC is intended rather for comparative analysis of several models and does not contain any information about the sensitivity and specificity of the model. According to the expert scale of AUC values accepted in the literature, which can be used to draw conclusions about the quality of the model, a very high quality model was built. The ideal model has 100% sensitivity and specificity (Pandey, 2019). However, in practice it is impossible to achieve this, nor is it possible to simultaneously increase the sensitivity and specificity of the model.

Using the obtained binary logistic regression model, it is possible to predict the probability of criminal recidivism separately for a particular convict on the basis of his/her individual values included in the model of variables.

**Conclusions**

We have discussed some aspects of the economics of incarceration, in particular, the possibility of reducing the economic costs of keeping prisoners in penitentiary institutions through the introduction of probation and possible parole. The results of the study proved the connection of the propensity of convicts to criminal recidivism with the following individual characteristics: age at the time of the first conviction (to the actual degree of punishment), type of employment at the time of conviction (up to actual punishment), number of conditional convictions, number of minor crimes, number of crimes of medium gravity and availability of granting early release.

The developed model allows to predict the probability of committing criminal recidivism on the basis of a persons individual statistical and dynamic data. The use of regression logistics models using up-to-date information on convicts is of great practical importance for making post-trial decisions, such as the grounds for possible parole; to support optimal pre-trial decisions, such as determining the amount of bail or the length of a sentence. Such predictive models can be used as a tool to assist law enforcement in carrying out crime prevention activities.

## References

Bąk, I., Wawrzyniak, K., Sobolewski, A. (2020). Classification of District Employment Agencies in Terms of Employment and Cost-Effectiveness Using Regression Trees. *Folia Oeconomica Stetinensia*, *20* (2), 20–35. DOI: 10.2478/foli-2020-0033.

Berezka, K.M., Berezsky, O.M., Masliy, V.V. (2013). Assessment of regional disparities of foreign investments in Ukraine. *Actual Problems of Economics*, *150* (12), 106–114.

Chirakijja, J. (2018). The Local Economic Impacts of Prisons. *Department of Econometrics and Business Statistics Monash University*. Retrieved from https://economics.smu.edu.sg.

Classification: ROC Curve and AUC (2022) *Google Developers*. Retrieved from https://developers.google.com.

Clay, A., Phillips, K.M. (2013) The Economy of Punishment. *Hurvard Business Review*. Retrieved from https://hbr.org.

Grawert, A., Craigie, T.-A. (2020). Mass. Incarceration Has Been a Driving Force of Economic Inequality. *Brennar Center for Justice*. Retrieved from https://www.brennancenter.org.

Dmytrów, K., Gdakowicz, A., Putek-Szeląg, E. (2019). Statistical Relations of the Qualitative Attributes of Real Properties Subject to Mass Appraisal. *Folia Oeconomica Stetinensia*, *19* (2), 25–37. DOI: 10.2478/foli-2019-0011.

Economics of Incarceration (2021). *Prison Policy Initiative*. Retrieved from https://www.prisonpolicy.org.

Kovalchuk, O. (2020). Smart Technologies in Criminal Justice. *International Scientific Conference "Iformation Tehnologies and Computer Modelling"*.

Kovalchuk, O., Shynkaryk, M., Masonkova, M. (2021). Econometric models for estimating the financial effect of cybercrimes. *11th International Conference "Advanced Computer Information Techno-logies"*. DOI: 10.1109/ACIT52158.2021.9548490.

McLaughlin, M., Pettus-Davis, C., Brown, D., Veeh, C., Renn T. (2016). The Economic Burden of Incarceration in the United States. *Institute for justice research and development*. Retrieved from https://ijrd.csw.fsu.edu.

Menard, S.W. (2002). *Applied Logistic Regression Analysis* (2nd ed.). London: New Delhi.

Mumford, M., Whitmore Schanzenbach, D., Nunn, R. (2020). *The economics of private prison. REPORT*. Brookings. Retrieved from https://www.brookings.edu.

Pandey, P. (2019). Simplifying the ROC and AUC metrics. *Towardsdatascience*. Retrieved from https://towardsdatascience.com.

Pilch, B. (2021). An analysis of the effectiveness of bankruptcy prediction models – an industry approach. *Folia Oeconomica Stetinensia*, *21* (2), 76–96. DOI: 10.2478/foli-2021-0017.

Stanimir, A. (2020a). Methods of a Multivariate Analysis of Non-Metric Data in Evaluating the Generational Perception of Social Characteristics. *Folia Oeconomica Stetinensia*, *20* (1), 390–407. DOI: 10.2478/foli-2020-0023.

Stanimir, A. (2020b). The Perception of the Generational Assessment of Selected Social Behaviour – A Confirmatory Factor Analysis. *Folia Oeconomica Stetinensia*, *20* (2), 346–360. DOI: 10.2478/foli-2020-0052.

Szkutnik, T. (2020). Identification of Outliers in High Density Areas with the Use of a Quantile Regression Model. *Folia Oeconomica Stetinensia*, *20* (2), 375–391. DOI: 10.2478/foli-2020-0054.

Vysotska, O., Grechanyuk S., Nesin I. (2022). In search of the optimal proportion… or a few facts about the closure of institutions. *State Penitentiary Service of Ukraine*. Retrieved from https://kvs.gov.ua.

*World Prison Brief* (2021, Dec.). Institute for Crime & Justice Policy Research. Retrieved from https://www.prisonstudies.org.

## Citation

Berezka, K.M., Kovalchuk, O.Ya., Banakh, S.V., Zlyvko, S.V., Hrechaniuk, R. (2022). A Binary Logistic Regression Model for Support Decision Making in Criminal Justice. *Folia Oeconomica Stetinensia*, *22* (1), 1–17. DOI: 10.2478/foli-2022-0001.