

Recognizing the Fictitious Business Entity on Logistic Regression Base

Andriy Krysovaty^a, Hrystyna Lipianina-Honcharenko^a, Svitlana Sachenko^a, Oksana Desyatnyuk^a, Arkadiusz Banasik^b and Iryna Lukasevych-Krutnyk^a

¹ West Ukrainian National University, Lvivska Str., 11, Ternopil, 46000, Ukraine

² Silesian University of Technology, Kaszubska Str., 23, Gliwice, 44-100, Poland

Abstract

The mechanism of creation and further activity of fictitious enterprises is constantly improved, and this one requires the use of adequate means to combat them. The method of fictitious enterprise detection on the basis of machine learning, namely Logistic Regression is offered. The developed method is represented by an algorithm and implemented in the software R environment. For example, a binary sample of 1048 enterprises was selected, of which 390 were fictitious. An EDA analysis was performed, which makes it possible to analyze the set and perform data cleaning if necessary. The correlation diagram shows that most of the parameters are slightly correlated with each other, only partially correlated with the parameter K205.

Keywords

fictitious enterprises, business entities, classification, machine learning, Support Vector Machine Classification.

1. Introduction

Shadow commodity-money transactions carried out using fictitious entrepreneurship are of concern in the economy of any country. Components of this criminal activity are commercial banks, a network of fictitious enterprises, legal enterprises. At the same time, fictitious entrepreneurship is an instrument of committing a number of mercenary crimes, in particular, tax evasion, smuggling, fraud with financial resources, etc. The spread of fictitious entrepreneurship is due to various reasons. For example, in Ukraine this has happened due to the long-term legal unregulation of private property, market and other social relations in the past and the imperfection of the state and legal mechanism for regulating business activities in modern conditions. Every year, the fiscal authorities of Ukraine alone identify about 6,000 fictitious legal entities. Given that the average turnover through the accounts of each of these categories of enterprises is about 5 billion dollars, budget losses from VAT alone amount to about 200 million USD dollars annually [21].

According to OLAF in 2020 (the year of the fight against COVID-19), 230 investigations were completed, 375 recommendations were issued to relevant national and European authorities, € 293.4 million was recommended for recovery in the EU budget and 290 new studies were opened after 1,098 preliminary analyzes. conducted by OLAF experts [26].

Defining an economic crime is a rather time-consuming procedure for law enforcement officers. Therefore, the development of an effective approach to identifying a fictitious enterprise is relevant.

IntellTSIS'2022: 3rd International Workshop on Intelligent Information Technologies and Systems of Information Security, March 23–25, 2022, Khmelnytskyi, Ukraine

EMAIL: rektor@wunu.edu.ua (A. Krysovaty); xrustya.com@gmail.com (H. Lipianina-Honcharenko); s_sachenko@yahoo.com (S. Sachenko); o.desyatnyuk@wunu.edu.ua (O. Desyatnyuk); arkadiusz.banasik@gmail.com (A. Banasik); lukru@ukr.net (I. Lukasevych-Krutnyk)

ORCID: 0000-0002-5850-8224 (A. Krysovaty); 0000-0002-2441-6292 (H. Lipianina-Honcharenko); 0000-0001-8225-1820 (S. Sachenko); 0000-0002-1384-4240 (O. Desyatnyuk); 0000-0002-4267-2783 (A. Banasik); 0000-0002-9557-7886 (I. Lukasevych-Krutnyk).



© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

This paper proposes a method for detecting a fictitious enterprise based on classical machine learning, namely Logistic Regression, the structure of which is as follows. Section 2 discusses the analysis of related work; Section 3 presents the method of classification of fictitious enterprises based on Logistic Regression, and Section 4 the implementation of the algorithm itself. Section 5 presents the conclusions of the study.

2. Related Work

Classification methods based on logistic regression are used in many areas: classification of tweet authors [4]; detection of fake news on social networks [5]; classification of diseases [6]; bankruptcy forecasting [7]; classification of production processes [8]; classification of text and images [9, 24]; in tourism [10]; cash flow forecasting [20].

In the article [29, 30] the research in the decision of the supposed good and unintentional bad consequences of application of artificial intelligence in financial crimes is carried out. In [1] a new machine learning system for adult and adolescent autism screening has been proposed, which contains vital features and performs predictive analysis using logistic regression to identify important information related to autism screening. Article [2] proposes a model for predicting the dismissal of employees based on machine learning. The prediction model is implemented using logistic regression using the cross-entropy function as the objective function and using Newton's method and regularization to optimize the model. In order to achieve the best, the logical regression of the kernel based on the confusion matrix (CM-KLOGR) is also proposed in [3].

The article [15] discusses the latest economic research on tax compliance and their application. Based on a unique set of data [13] on the leakage of client lists from offshore financial institutions, it was found that tax evasion is very concentrated among the rich. The document [14] uses administrative microdata to study the impact of law enforcement efforts on taxpayers' reporting of offshore accounts and revenues. Increasing the exchange of information between countries is a key policy tool in the fight against cross-border tax evasion, in [16] the short-term effect of the Common Reporting Standard (CRS). CRS is the first global multilateral standard for automatic exchange of information. In [12], a model of tax evasion was developed. In [18], a model is proposed to study the relationship between economic growth and both types of income tax evasion.

Article [19] examines the issues of tax avoidance and the evolution of tax evasion, highlighting the factors that influence the emergence of these phenomena from a historical point of view: it is determined who is a typical fraudster, as auditors can identify problems with tax evasion and avoidance of pay taxes because they know better the type of person who can commit tax fraud, as this can be seen as an element of risk in the audit. To combat tax evasion, the OECD has developed an automatic information exchange (AIE) standard, in [17], the factors that explain the differences between the two information exchange mechanisms in the implementation of the AIE standard. The results of the study show that the differences are influenced by existing IT capabilities, compatibility, trust between information exchange partners, differences in power, inter-organizational relations and the expected benefits of implementing such mechanisms. Article [28] examines the processes of money laundering or, more broadly, illegal financial transactions, such as terrorist financing.

It should be noted that the above-mentioned works do not describe the detection of fictitious enterprises with the help of information technology. On the other hand, the closest analogues [1-3], representing the Logistic Regression classification, do not investigate the use of this method to determine a fictitious enterprise.

Thus, the purpose of this article is to develop a method of classification of fictitious enterprises on the basis of Logistic Regression as a basis for the appropriate software environment.

3. Materials and Methods

To be able to quickly identify a fictitious enterprise, a method of classification of fictitious enterprises based on machine learning by the Logistic Regression method has been developed. The advantages of the logistic regression method are: well studied; very fast, can work on very large samples; practically out of competition, when the signs are very many (from hundreds of thousands and

more), and they are sparse; the coefficients before the signs can be interpreted; gives the probability of assignment to different classes. The disadvantages of this method are: they work poorly in problems in which the dependence of the answers on the signs is complex, nonlinear.

The developed method is represented by an algorithm (Fig. 1) and the following steps.

Step 1. At the beginning the data are entered: enterprise code (ID); parameter for determining the fictitiousness of the enterprise (Fit); company name (Company); legal address (Address); physical address (FAddress1,... FAddressn); KVED; names of managers (PIPKER); photo equipment with geolocation (Foto); availability of a register of legal entities and individuals (EDR) in a single database; availability in the database of VAT payers (P); timely payment of taxes (PO); availability of settlements with co-agents (K); information on the presence of company executives in the state register of declarations (VKK); availability of licenses according to NACE (L); the presence of criminal cases under Art. 205 of the Criminal Code of Ukraine (K205); presence of mentions of company executives with keywords: criminal case, corruption, offshore accounts, etc. (ZMI); availability of land at the legal or physical address (ZD); availability of registered trademarks and services, database of industrial marks, database of inventions and other databases of the Institute of Industrial Property of Ukraine (TovZ); availability of issued motor third party insurance policies, MTIBU policy check, motor third party database, search by state car number, check of the status of the Green Card policy for cars owned by the company (SP); availability of cars and their owners issued to the company (A); coincidence of registered cars with insurance policies (A&SP); availability in the database of exporters (E); availability in the stock market database (F); the presence of cars and their owners registered with the company wanted (AR); the presence of weapons of the owners of the company wanted (ZR); the presence of cultural values of the owners of the company wanted (KR); availability of construction licenses in the company (LB); availability of real estate in the company (NM); availability of the company's website (NS); availability of equipment, recognition of equipment by the available photo and determination of compliance of geolocation with the production address (FR); availability of the company's social networks and affiliated employees (FC).

Step 2. Conducting an EDA with the output of the results. Intelligence data analysis (EDA) – analysis of the basic properties of data, finding in them general patterns, distributions and anomalies, construction of initial models, often using visualization tools.

Step 3. Data conversion to binary expression.

Step 4. Data cleaning [22, 25].

Step 5. Data distribution. The data are divided into 25% of the test and 75% of the training sample.

Step 6. Select the variable most correlated with Fit [22, 23].

Step 7. Then the model, which contains only one selected independent variable, is checked for significance using a private F-test. If the significance of the model is not confirmed, the algorithm ends because of the lack of significant input variables. Otherwise, this variable is entered into the model and the transition to the next point of the algorithm.

Step 7.1. For the remaining variables according to formula (1), the value of the statistics μ is calculated, which is the ratio of the increase in the sum of regression squares achieved by introducing the corresponding additional variable into the model to the value $MSE^{<full>}$.

$$h_{\beta}(x) = g(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) = g(\beta^T x), \quad (1)$$

where $g(z) = \frac{1}{1+e^{-z}}$

$$MSE^{<full>} = \frac{SSE}{df^{<SSE>}} = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-k-2} \quad (2)$$

where SSE is the sum of the squares of errors (the model is constructed on variables), accounting for one degree of freedom ($X^{<1>}, X^{<2>}, \dots, X^{<k>}$ and $X^{<extra>}$ $df^{<sse>}$).

Step 7.2. Calculated F_{real}

$$F_{real} = \frac{SSE - 0}{MSE^{<full>}}$$

Step 8. There is a comparison F_{real} with F_{table} , which indicates the need to include the variable X_n in the regression model, while the probability that the decision to include will be incorrect is $\alpha = 0.05$ (the values are taken from the Fisher criterion table) [11].

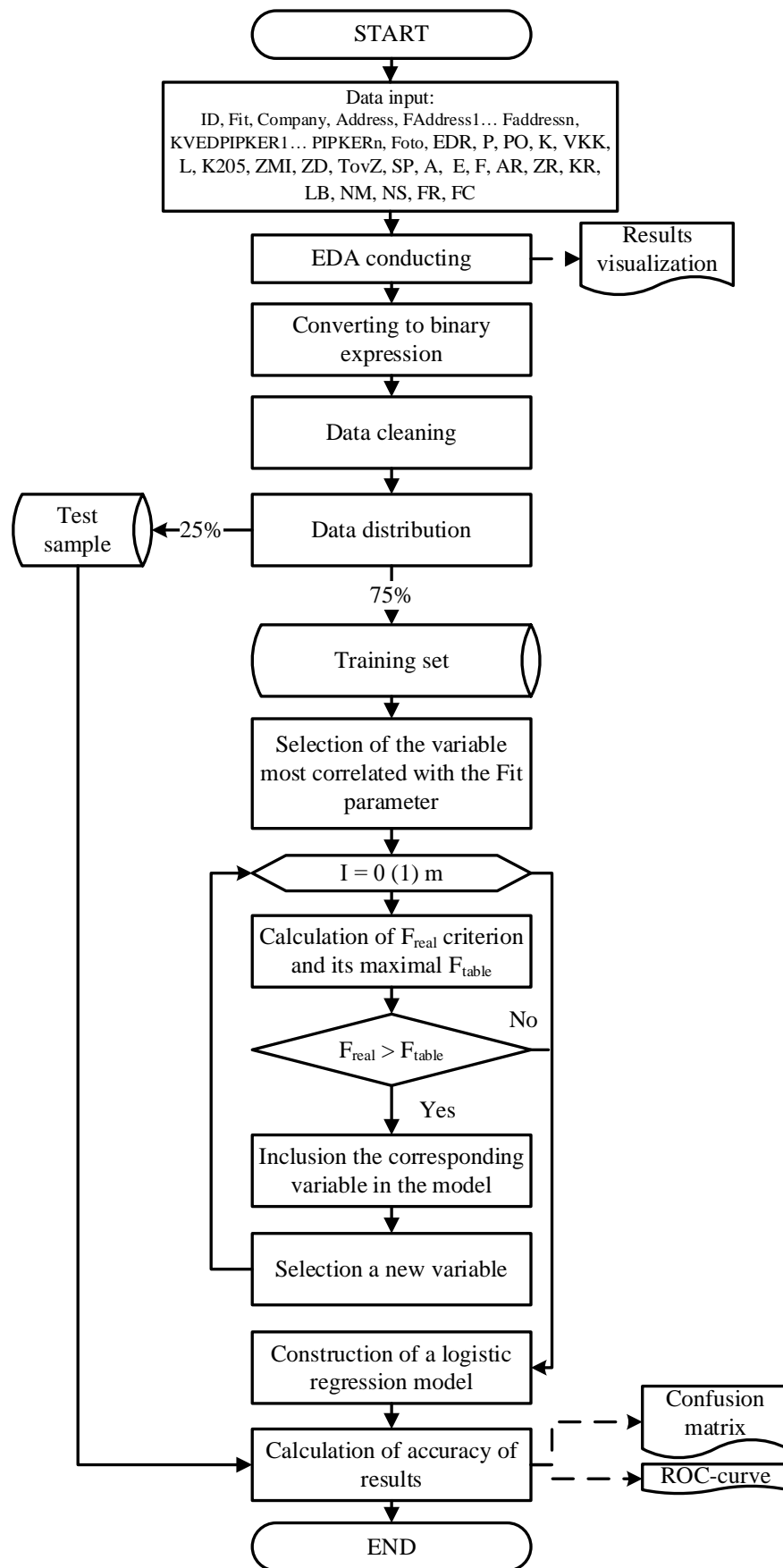


Figure 1: Algorithm for classification of fictitious enterprises on the basis of machine learning by the method of Logistic Regression

Step 9. From all variables-applicants for inclusion in the model, the one that has the highest value of the criterion calculated in step 8 is selected.

Step 10. The significance of the independent variable selected in step 9 is checked. If its significance is confirmed, it is included in the model, and the transition to step 8 (but with a new independent variable in the model). Otherwise, the algorithm stops.

Step 11. The construction of a logistics model based on the obtained variables in step 9.

Step 12. Based on the test sample, the obtained model is tested for the output of results, namely Confusion matrix and ROC-analysis.

4. Experimental Results and Discussion

To implement the algorithm (Fig. 1) of fictitious enterprises classification based on machine learning by Logistic Regression, selected free programming R language. R language allows to model statistical indicators, having a large number of relevant libraries and easy to operate, so it is a very good choice for this task.

A sample of 1,048 enterprises was selected to solve this problem. 390 of them are fictitious. Next, an EDA analysis was performed, namely, the distribution of the fit parameter is presented. The visualized information `plot_normality()` is as follows: Histogram of original data; Q-Q plot of original data; histogram of log transformed data; Histogram of square root transformed data. From the results, the binary values are clearly traced.

For binary values, the main graphical result is the correlation matrix (Fig. 2). The diagram shows that most of the parameters are poorly correlated with each other, only partially traced correlation with the parameter K205, namely with the parameter indicating the presence of criminal cases under Art. 205 of the Criminal Code of Ukraine. Due to the fact that there are small traces of correlation, it is difficult to determine whether the company belongs to the fictitious or not, so it is the classification algorithms based on machine learning, ideal for the task.

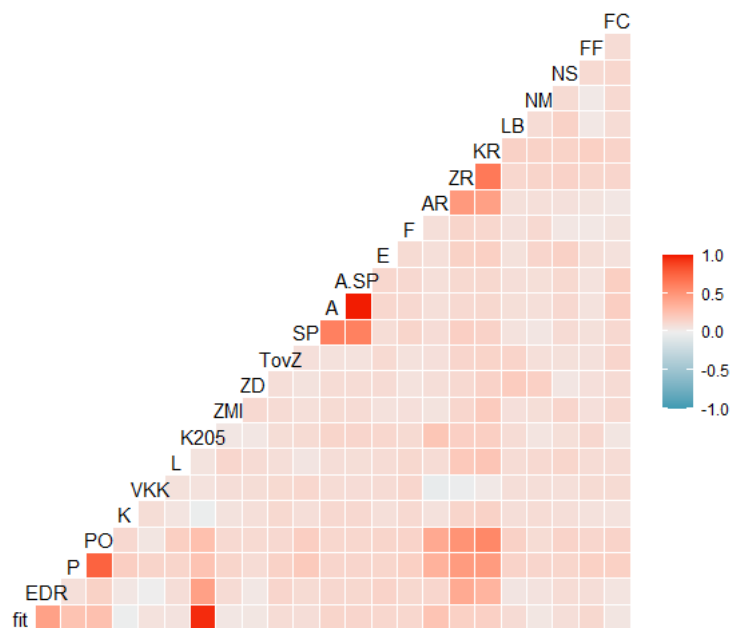


Figure 2: Correlation

For testing, the data set was divided into a training set (75%) and a testing set (25%). We train the model to predict a fictitious enterprise. Instead of directly modeling the answer Y, logistic regression simulates the probability that Y belongs to a certain category, in our case, the probability of fictitiousness. This probability can be calculated using the logistics function. Thus, we build a model based on logistic regression (Fig. 3).

```

Call:
glm(formula = fit ~ ., family = binomial(link = "logit"), data = trainset)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.5873 -0.1007 -0.0047  0.1050  2.0492

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -13.869554  1.662398  -8.343  <2e-16 ***
ID           0.020020  0.001818  11.012  <2e-16 ***
EDR          0.837101  1.328223   0.630  0.529
P            0.010901  1.139985   0.010  0.992
PO           0.130605  0.826281   1.368  0.171
K           -0.278529  0.371038  -0.751  0.453
VKK          0.056051  0.355929   0.157  0.875
L            0.247243  0.345017   0.717  0.474
K205         0.946000  0.993254   0.952  0.341
ZMI          -0.102556  0.337747  -0.304  0.761
ZD           0.527000  0.350854   1.502  0.133
TovZ        -0.513267  0.346379  -1.482  0.138
SP           0.423517  0.426886   0.992  0.321
A           -0.653820  0.510066  -1.282  0.200
`A&SP`      NA         NA         NA     NA
E           -0.232298  0.342067  -0.679  0.497
F           -0.080450  0.356129  -0.226  0.821
AR           0.182801  0.949578   0.193  0.847
ZR           0.321103  1.040714   0.309  0.758
KR           0.086097  0.832742   0.103  0.918
LB          -0.448565  0.343694  -1.305  0.192
NM          -0.062249  0.352855  -0.176  0.860
NS           0.503829  0.345578   1.458  0.145
FF           0.483070  0.352385   1.371  0.170
FC          -0.426550  0.336234  -1.269  0.205
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1087.16  on 785  degrees of freedom
Residual deviance:  243.54  on 762  degrees of freedom
AIC: 291.54

Number of Fisher Scoring iterations: 8

```

Figure 3: Construction of the model

As can be seen, from the obtained results, standard errors, z-score and p-values for each of the coefficients were determined. None of the coefficients are significant here, except for K205 and EDR, which is similarly represented by correlation (see Fig. 2). The effectiveness of logistic regression is assessed by certain key indicators:

- AIC (Akaike Information Criteria): this is the equivalent of R2 in logistic regression. It measures suitability when a fine is applied to a number of parameters. Smaller AIC values indicate that the model is closer to the truth. In the presented implementation, AIC = 291.54.
- Zero deviation: suitable for model with interception only. Degree of freedom n-1. Interpreted as a Chi-square value (an adapted value that differs from the actual value hypothesis test). Residual deviations: model with all variables. This is also interpreted as a test of the chi-square hypothesis. The example shows (see Fig.3) that the deviation decreases by 843.62 when subtracting 23 variables of the predictor (degree of freedom = number of observations – the number of predictors). This reduction in deviation is evidence of the suitability of the obtained model.
- Number of iterations estimated by Fisher: the number of iterations before convergence, equal to 8, for the task.

Now let's see how accuracy, sensitivity and specificity are transformed for a given threshold. By default, use the 50% threshold to determine the probability of fictitiousness to assign class observations. However, from the graph (Fig. 4), it is seen that the probability threshold has two increases from 1% to 50% and 50% to 100%.

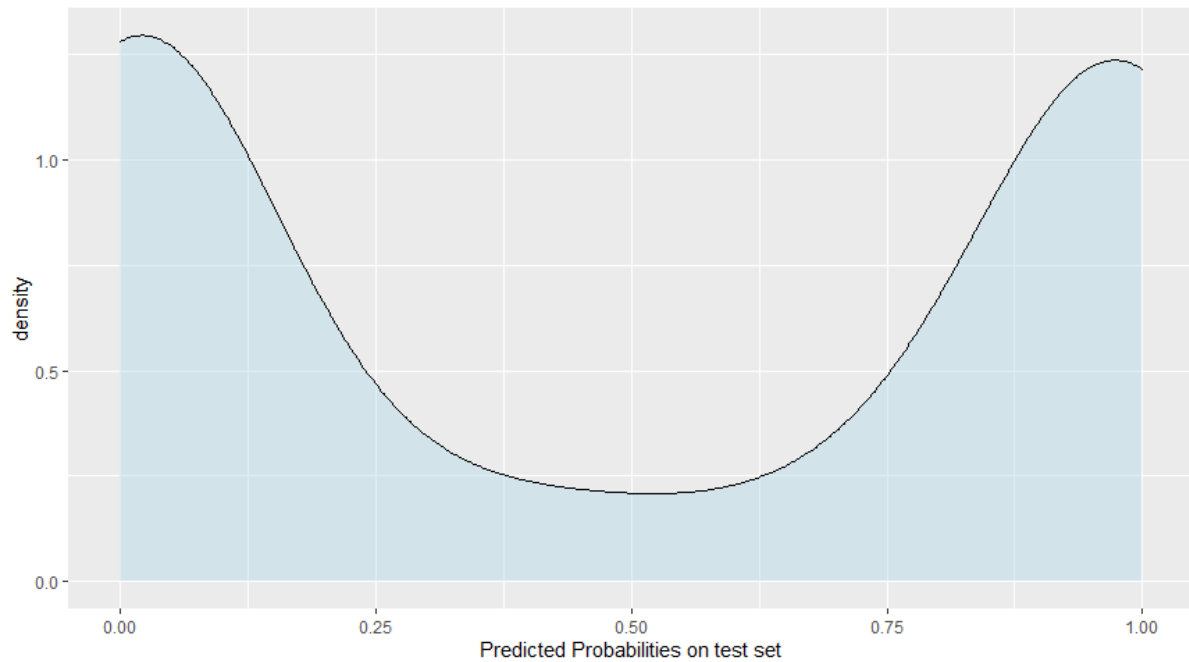


Figure 4: Predicted Probabilities on test set

Let's consider the indicators of accuracy, sensitivity and specificity (Fig. 5), the diagram shows that the accuracy and sensitivity, begins to decrease at 55%. Therefore, consider the confusion matrix (Fig. 6) for the cut-off point by 55%, with Accuracy: 0.99.

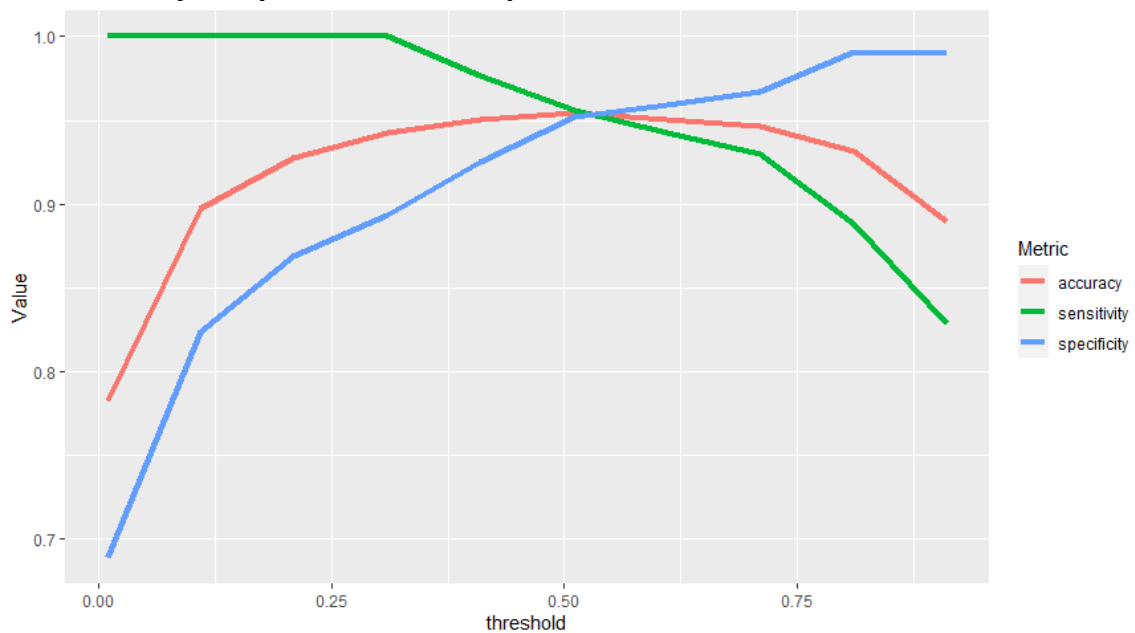


Figure 5: Analysis of accuracy, sensitivity and specification of the obtained model

Table 1

Confusion matrix

		Actual	
		0	1
Predicted	0	129	5
	1	7	121

The different values of the Confusion matrix (Table 1) will be as follows for the training sample:

- True positive (TP) = 129; this means that 129 indicators of positive class data are correctly classified by the model;
- True negative (TN) = 121; this means that 121 data points of negative class were correctly classified by the model;
- False positive (FP) = 5; this means that 5 indicators of negative class data were incorrectly classified as models belonging to the positive class;
- False negative (FN) = 7; this means that 7 data indicators of the positive class were incorrectly classified as models belonging to the negative class.

The ROC curve is a popular graph for displaying two types of errors simultaneously for all possible thresholds. Therefore, we present the ROC-curve for our study (Fig. 6).

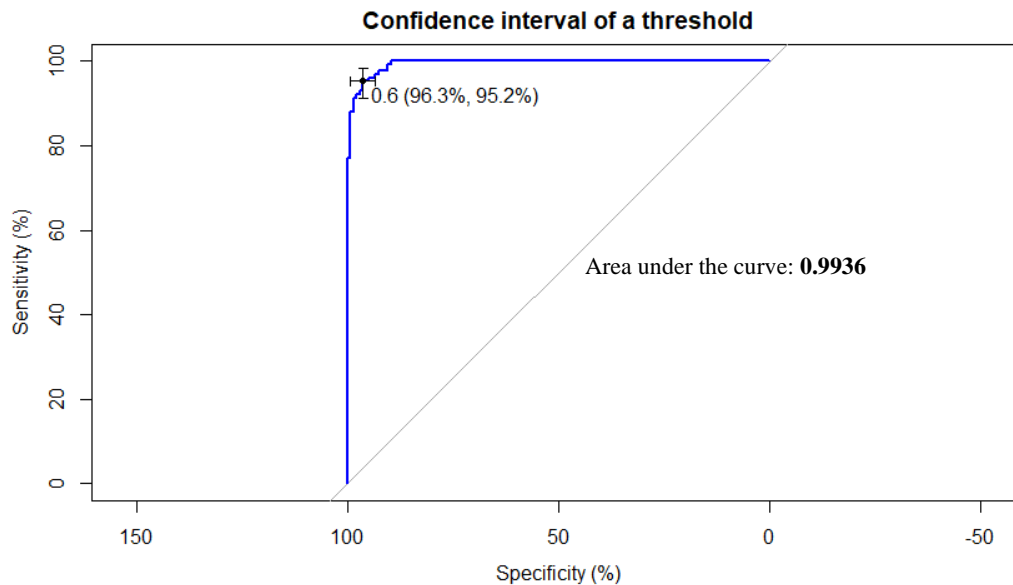


Figure 6: Confidence interval of a threshold

As shown in the ROC curve (see Fig. 6), the optimal threshold level of diagnostic assessment for forecasting fictitious enterprises is 0.6, sensitivity and specificity are 96.3% and 95.2%, respectively. The forecast for predicting the accuracy of determining fictitious enterprises is quite high, namely $AUC = 0.99$.

The logistic regression model was used to predict the fictitiousness of the enterprise. Clipping 55% gave a high Accuracy: 0.99, and the area curve also provides the same accuracy of 0.99.

5. Conclusions

A method of detecting a fictitious enterprise based on the classic method of machine learning, namely Logistic Regression, is proposed, which allows to quickly track fictitious enterprises, which is useful for public sector employees to prevent economic crimes.

The method is implemented in the software environment R. To solve this problem, a binary sample of 1048 enterprises was selected, of which 390 are fictitious. The EDA analysis allows to clear the data as needed. The correlation diagram shows that most parameters are poorly correlated with each other. In particular, there is only a partial correlation with parameter K205, namely with the parameter indicating the existence of criminal cases under the Criminal Code of Ukraine. Logistic Regression model built: $AIC = 291.54$; the deviation decreases by 843.62 when subtracting 23 predictor variables; number of iterations according to Fisher = 8. Prediction of fictitiousness of enterprises is carried out: Accuracy = 0.99; $AUC = 0.99$. The Confusion matrix derived the following classification results for the training sample: 129 indicators of positive class data correctly classified by the model; 121 data points of negative class were correctly classified by the model; 5 indicators of negative class data were incorrectly classified as models belonging to the positive class; 7 indicators of positive class data were incorrectly classified as models belonging to the negative class.

In further research, it is expected to develop an algorithm for recognizing images of enterprises equipment with geolocation data processing and converting them into binary values.

6. References

- [1] F. Thabtah, N. Abdelhamid, & D. Peebles, A machine learning autism classification based on logistic regression analysis. Springer. *Health Inf Sci Syst* 7 (2019) article ID 12. <https://doi.org/10.1007/s13755-019-0073-5>
- [2] W. Dai, Z. Zhu, Employee resignation prediction model based on machine learning. In: Abawajy J., Choo KK., Xu Z., Atiquzzaman M. (eds), *Proceedings of the 2020 International Conference on Applications and Techniques in Cyber Intelligence (ATCI'2020), Advances in Intelligent Systems and Computing*, 1244, (2020) 367-374. https://doi.org/10.1007/978-3-030-53980-1_55
- [3] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe and A. Ralescu, Confusion-matrix-based kernel logistic regression for imbalanced data classification. *IEEE Transactions on Knowledge and Data Engineering* 29.9 (2017) 1806-1819, doi: 10.1109/TKDE.2017.2682249.
- [4] O. Aborisade and M. Anwar, Classification for authorship of tweets by comparing logistic regression and Naive Bayes Classifiers, in: *Proceedings of the 2018 IEEE International Conference on Information Reuse and Integration (IRI)*, (2018) 269-276, doi: 10.1109/IRI.2018.00049.
- [5] M. Goksu, N. Cavus, Fake news detection on social networks with artificial intelligence tools: Systematic literature review. in: Aliev R., Kacprzyk J., Pedrycz W., Jamshidi M., Babanli M., Sadikoglu F. (eds) *10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions ICSCCW-2019. Advances in Intelligent Systems and Computing*, 1095 (2020) 47-53. https://doi.org/10.1007/978-3-030-35249-3_5
- [6] L. Liu, Research on logistic regression algorithm of breast cancer diagnose data by machine learning, in: *Proceedings of the 2018 International Conference on Robots & Intelligent System (ICRIS)*, (2018) 157-160, doi: 10.1109/ICRIS.2018.00049.
- [7] F. Barboza, H. Kimura, E. Altman, Machine learning models and bankruptcy prediction. *Expert Systems with Applications* 83 (2017) 405-417, doi: 10.1016/j.eswa.2017.04.006.
- [8] İ. Kabasakal, F.D. Keskin, A. Koçak, H. Soyuer, A prediction model for fault detection in molding process based on logistic regression technique. In: Durakbasa N., Gençyılmaz M. (eds) *Proceedings of the International Symposium for Production Research ISPR'2019, Lecture Notes in Mechanical Engineering*. (2020) 351-360. https://doi.org/10.1007/978-3-030-31343-2_31
- [9] M. Nieuwenhuis, & J. Wilkens, Twitter text and image gender classification with a logistic regression n-gram model. in: *Proceedings of the Ninth International Conference of the Working Notes of CLEF 2018 – Conference and Labs of the Evaluation Forum (CLEF 2018)*. CEUR-WS, 2125.
- [10] V. Krylov, A. Sachenko, P. Strubyskyi, D. Lendiuk, H. Lipyanina, D. Zahorodnia, V. Dorosh, & T. Lendyuk, Multiple regression method for analyzing the tourist demand considering the influence factors. in: *Proceedings of the 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS'2019)*, Metz, France, 2 (2019) 974-979.
- [11] Upper Critical Values of the F Distribution. Information Technology Laboratory | NIST. URL: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda3673.htm>
- [12] M. G. Allingham, A. Sandmo, Income tax evasion: a theoretical analysis. *Journal of Public Economics* 1 (1972) 323–338, doi:10.1016/0047-2727(72)90010-2.
- [13] A. Alstadsæter, N. Johannesen, and G. Zucman, Tax Evasion and Inequality. *American Economic Review* 109 (2019) 2073-2103, Doi: 10.1257/aer.20172043.
- [14] N. Johannesen, P. Langetieg, D. Reck, M. Risch & J. Slemrod, Taxing hidden wealth: The consequences of US enforcement initiatives on evasive foreign accounts. *American Economic Journal: Economic Policy* 12 (2020) 312-346.
- [15] J. Slemrod, Tax compliance and enforcement. *Journal of Economic Literature* 57 (2019) 904-954, doi: 10.1257/jel.20181437.
- [16] E. Casi, C. Spengel, & B. M. Stage, Cross-border tax evasion after the common reporting standard: Game over? *Journal of Public Economics* 190 (2020), 104240.

- [17] R. A. Kurnia, D. Praditya, M. Janssen, A comparative study of business-to-government information sharing arrangements for tax reporting. in: Dwivedi Y., Ayaburi E., Boateng R., Effah J. (eds) *ICT Unbounded, Social Impact of Bright ICT Adoption TDIT 2019*. IFIP Advances in Information and Communication Technology, 558 (2019) 154-169. https://doi.org/10.1007/978-3-030-20671-0_11
- [18] C. Bethencourt, L. Kunze, Social norms and economic growth in a model with labor and capital income tax evasion. *Economic Modelling* 86 (2019) 170-182. doi: 10.1016/j.econmod.2019.06.009.
- [19] D. Saxunova, R. Sulikova, R. Szarkova, Tax management hierarchy – Tax fraud and a fraudster. in: *Proceedings of the Joint International Conference on Managing the Global Economy MIC'2017, Monastier di Treviso, Italy, 24–27 May 2017, University of Primorska Press*.
- [20] K. Bazilevych, M. Mazorchuk, Y. Parfeniuk, V. Dobriak, I. Menailov, & D. Chumachenko, Stochastic modelling of cash flow for personal insurance fund using the cloud data storage. *International Journal of Computing*, 17 (2018) 153-162. <https://doi.org/10.47839/ijc.17.3.1035>
- [21] UNITED NATIONS DEPARTMENT FOR ECONOMIC AND SOCIAL AFFAIRS. (2020). *World economic situation and prospects 2020*. p. 236.
- [22] H. Lipyana, S. Sachenko, T. Lendyuk, V. Brych, V. Yatskiv, O. Osolinskiy Method of Detecting a Fictitious Company on the Machine Learning Base. In: Hu Z., Petoukhov S., Dychka I., He M. (eds) *Advances in Computer Science for Engineering and Education IV. ICCSEE 2021. Lecture Notes on Data Engineering and Communications Technologies*, 83 (2021) doi: https://doi.org/10.1007/978-3-030-80472-5_12
- [23] A. Krysovaty, H. Lipyana-Goncharenko, S. Sachenko and O. Desyatnyuk. Economic Crime Detection Using Support Vector Machine Classification. *Modern Machine Learning Technologies and Data Science Workshop. Proc. 3rd International Workshop (MoMLeT&DS 2021)*. Volume I: Main Conference. Lviv-Shatsk, Ukraine, June 5-6, 2021, 830-840.
- [24] R. Gramyak, H. Lipyana-Goncharenko, A. Sachenko, T. Lendyuk and D. Zahorodnia. Intelligent Method of a Competitive Product Choosing based on the Emotional Feedbacks Coloring, *CEUR WS*, (2021) 246-257.
- [25] Z. Hu, M. Ivashchenko, L. Lyushenko, D. Klyushnyk, Artificial Neural Network Training Criterion Formulation Using Error Continuous Domain, *International Journal of Modern Education and Computer Science (IJMECS)*, 13 3 (2021) 13-22. doi: 10.5815/ijmeecs.2021.03.02
- [26] The OLAF report 2020. OLAF, 2020, URL: https://ec.europa.eu/anti-fraud/system/files/2021-12/olaf_report_2020_en.pdf
- [27] M. Kantardzic, *Data mining: concepts, models, methods, and algorithms*. 3rd Edition. Wiley-IEEE Press. 2019, 672 p.
- [28] L. Corselli, "Italy: money transfer, money laundering and intermediary liability", *Journal of Financial Crime*, (2020) Vol. ahead-of-print No. ahead-of-print. doi: <https://doi.org/10.1108/JFC-10-2019-0137>
- [29] P. Yeoh, Artificial intelligence: accelerator or panacea for financial crime?", *Journal of Financial Crime*, 26 2 (2019) 634-646. doi: <https://doi.org/10.1108/JFC-08-2018-0077>
- [30] S. Dsouza, H. Habibniya, R. Demiraj, AI, a Provenance or Solution for Financial Crime. *Manag Econ Res J*, 7(2) (2021) 26140.