APPLICATION OF ARTIFICIAL INTELLIGENCE TO SUPPORT LAW ENFORCEMENT AGENCIES IN COMBATING FINANCIAL CRIME

Paweł Olber[1]
Police Academy in Szczytno, Poland
Wojciech Warczak[2]
Police Academy in Szczytno, Poland

Abstract: The relentless technological development has far-reaching implications for almost every area of life. Fast and unlimited access to information, the remote establishment of interpersonal, business, and financial relations, including rapid money transfers, are only some developments. The development of new technologies, apart from many advantages, brings with it many risks. These threats include criminal activity on the Internet. It is aimed at obtaining financial gain. This type of criminal activity involves financial and cyber aspects, which causes many problems and challenges in law enforcement detection activities. Law enforcement activities must consider collecting and analysing vast amounts of data, which reach gigantic sizes and capacities. We describe an example of the application of artificial intelligence solutions in conducting financial investigations in this article. This work is a contribution to law enforcements agencies pragmatics. The publication shows the role and place and potential of artificial intelligence in combating financial crime.

Keywords: artificial intelligence, machine learning, economic crime, law enforcements agencies

INTRODUCTION

Rapid technological development characterises the social and economic development observed in Poland, Europe and worldwide. This is a sign of our times. Faster access to information, able to talk far a distance, the remote establishment of interpersonal, busi-

1 p.olber@wspol.edu.pl
2 w.warczak@wspol.edu.pl

ness and financial relations, including rapid money transfers, are just some manifestations of this development. Electronic devices designed to perform to our needs surround us, and (via the Internet) we can control these devices, even from the other side of the world. These solutions make life easier, but they also carry risks. We are operating on a virtual network, what means that other Internet users unknown to us may also have access to information about us. Their intentions will not always match ours, and they are often illegal. Criminals aim more often at obtaining financial gain. They often do this using the Internet. The preference of perpetrators for remote forms of relations with victims is not accidental and creates problems from the detection side.

From the point of view of law enforcement authorities, in this case, we are dealing with the combine of financial and cyber aspects of criminal activities. This has specific negative consequences for the effectiveness of the state in citizens and obliges to take measures aimed at solving the identified problematic situation.

This paper presents the possibility of using machine learning algorithms in automatic analysis of financial data sets, necessary from the point of view of combating economic crime. In this paper, we present an example of a financial dataset, which was subjected to preliminary analysis. The article presents the results of preliminary analytical activities and describes learning and evaluating the model together with optimisation. We will describe the results at the end of the article. The code used in the study is available on GitHub: https://git.io/JCgCK.

We first described the major challenges for law enforcement agencies in combating financial crime, highlighting the problem of analysing financial data sets. Then we described the sources and types of financial data, including bank data and sets of financial transactions. For this article, we have used a set of crafted data, which we have subjected to preliminary analysis. We then carried out supervised learning and evaluation of selected classifier models together with optimisation. Following the research part, we presented the results. We conclude the paper and discuss future research directions and existing limitations. The paper ends with conclusions.

MAJOR CHALLENGES FOR LAW ENFORCEMENT AGENCIES IN COMBATING FINANCIAL CRIME

Financial crime (committed in both the virtual and real worlds) mostly involves frauds, which accounts for the vast majority of economic crime (Ministry of Internal Affairs and Administration of the Republic of Poland, 2017). The manifestations of such activities are excessive, ranging from tax (treasury), financial, investment, para-banking, sales or access to service fraud, and their number and role are constantly increasing. Over 199,000 offences were recorded in Poland throughout 2020, up from 188,000 in 2019 and 181,000 in 2018, with a steady decline in the detection rate of offenders (Polish National Police, n.d.). It is not just the number of incidents that matter. Important is the increasing extent of organisation of the perpetrators. They use the latest technological solutions, as well as precise identification of the demand side (the most popular social needs), and thus the perfect adaptation of the used mechanisms to the victim (Nepelski & Struniawski, 2017). From the point of view of investigative institutions, this poses an increasing challenge, especially in establishing the mechanism of perpetration (modus operandi) and, above all, the identity

of the real beneficiaries of the crime (persons with significant control). Using modern technologies ensures anonymity of criminals and the rapid and remote transfer and hiding of funds. Thus, we are dealing with IT (technological) tools and with financial instruments. Law enforcement authorities intend to gather a full range of evidence in these types of proceedings. This has to aim at achieving the objectives of criminal proceedings: the detection of the perpetrator of a crime, the disclosure of the circumstances of the crime, and securing the interests of the injured party (Criminal Procedure Code of the Republic of Poland, 1997). In ongoing cases (operational and investigative) concerning financial crimes, law enforcements collect at least the following datasets:

- Call Data Records (telephone records, e-mail addresses, IP logs)

- financial data (bank accounts, information from payment service providers, information from telecommunications operators)

- open source data (from public databases and registers)

- closed data records (information and data got from entities and institutions within the framework of exercising rights to get information for conducted activities/proceedings).

Proceedings involving the activities of organised criminal groups (tax evasion, money laundering and larger-scale fraud, there are usually investigations related to organised crime groups), tax evasion, money laundering and fraud on a larger scale, proceedings usually last for several years and the amount of collected data sets and information (in paper and electronic form) reaches gigantic proportions and capacities (Leehealey & Chirgula, 2019). If we also consider the variety of formats of these data (telecommunications or financial operators are not obliged to make data available in a homogenous format), it is easy to conclude that the possibility for investigators to effectively analyse data in such cases, has long been exhausted. This may generate dangerous results, as the lack of support of law enforcement agencies with advanced technological tools may cause a further decrease in detection efficiency and ultimately a decrease in the level of public confidence in state structures and internal security (Ministry of Internal Affairs and Administration of the Republic of Poland, 2017).

SOURCES AND TYPES OF FINANCIAL DATA

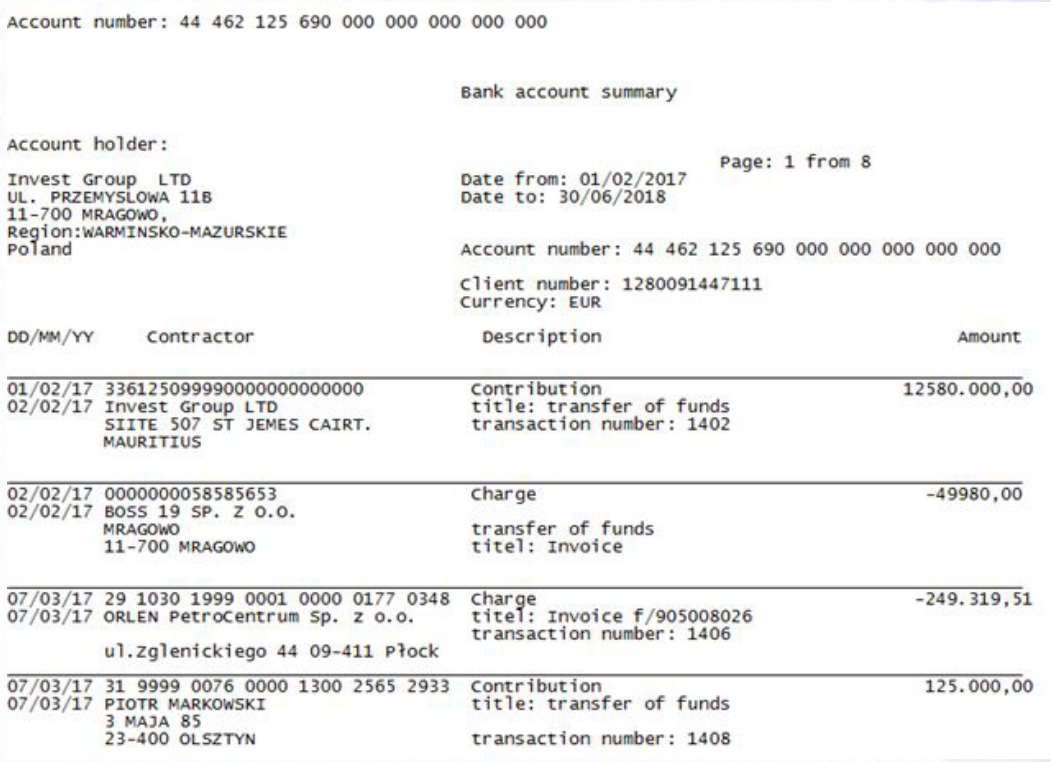
Financial data is a rich set which contains information on specific transactions, the sender of the payment, the addressee, the amount, the title, the time of order and the execution of the transfer, the currency, the transaction identifier, the addresses and entities). In a broader sense, financial data also include personal and company data (so-called registration data) related to the account holder and online data (related to logging banking and remote ordering of payments). Banks and other payment institutions collect these data in an electronic format, and then in justified cases transferred to allow institutions for criminal prosecution. According to information established by the authors based on data from the records of the Polish Financial Supervision Authority, the number of institutions, entities and companies which conduct official and legal business involving the process of financial data is over 2,000 in Poland (Polish Financial Supervision Authority, 2021).

We can divide the sources of financial data by the category of entities that collect and process them. In this context, we distinguish between:

- bank data,
- non-bank data (payment service providers, investment entities, lending companies).

BANKING DATA

A major obstacle to rapid analysis of bank data is its diverse format and quantity. Collecting this material to a common denominator is time-consuming and requires involvement and work of many officers. Financial data derived solely from the banking system may be in paper form or in various electronic formats in unstructured form. Below is an example of such bank data in text format.



Account number: 44 462 125 690 000 000 000 000 000

Bank account summary

Account holder:

Page: 1 from 8

Invest Group LTD
UL. PRZEMYSLOWA 11B
11-700 MRAGOWO,
Region:WARMINSKO-MAZURSKIE
Poland

Date from: 01/02/2017
Date to: 30/06/2018

Account number: 44 462 125 690 000 000 000 000 000

Client number: 1280091447111
Currency: EUR

DD/MM/YY	Contractor	Description	Amount
01/02/17	33612509999000000000000	Contribution	12580.000,00
02/02/17	Invest Group LTD SIITE 507 ST JEMES CAIRT. MAURITIUS	title: transfer of funds transaction number: 1402	
02/02/17	0000000058585653	Charge	-49980,00
02/02/17	BOSS 19 SP. Z O.O. MRAGOWO 11-700 MRAGOWO	transfer of funds titel: Invoice	
07/03/17	29 1030 1999 0001 0000 0177 0348	Charge	-249.319,51
07/03/17	ORLEN PetroCentrum Sp. z o.o. ul.Zglenickiego 44 09-411 Płock	titel: Invoice f/905008026 transaction number: 1406	
07/03/17	31 9999 0076 0000 1300 2565 2933	Contribution	125.000,00
07/03/17	PIOTR MARKOWSKI 3 MAJA 85 23-400 OLSZTYN	title: transfer of funds transaction number: 1408	

Figure 1. *Example of bank data in text format*

There are examples of cases of financial fraud, connected with money laundering of cross-border nature. These cases concerned several dozen bank accounts (including foreign ones) and contains data on several hundred thousand payment operations. In such circumstances, it becomes necessary to provide technological support for investigating authorities with tools for quick and efficient data analysis. The purpose of these tools may be to identify and block suspicious financial transactions.

DATASET OF FINANCIAL TRANSACTIONS

Showing the potential of artificial intelligence solutions to support law enforcement agencies in combating economic crime requires an appropriate dataset. Open collections lack information on real financial transactions. Therefore, for this article, the authors of the publication used a set of prepared data available through the Kaggle service. In selecting an appropriate dataset, the usefulness and transparency of the information guided the authors. Many of the datasets reviewed contained anonymised information, which did not allow for comprehend of their nature and the links and relationships that existed. Thus, given the purpose of this article, the authors selected a well-described dataset: Ecommerce Fraud Data (Rastogi, 2021). This set comprises two files: Customer_DF (1).csv and cust_transaction_details (1).csv, containing information about commercial transactions on the Internet. The authors conducted further research activities related to the dataset in question in the following order:

- data analysis and preparation;
- learning and evaluation of models;
- optimisation.

DATA ANALYSIS AND PREPARATION

We carried out the data analysis and other stages of the research activities through the Google Colab tool, which allows writing and run Python code in a web browser. To process the dataset, we used the libraries Pandas and NumPy, with which we read the structure of the files:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 623 entries, 0 to 622
Data columns (total 11 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Unnamed: 0                      623 non-null    int64
 1   customerEmail                   623 non-null    object
 2   transactionId                   623 non-null    object
 3   orderId                         623 non-null    object
 4   paymentMethodId                 623 non-null    object
 5   paymentMethodRegistrationFailure 623 non-null   int64
 6   paymentMethodType               623 non-null    object
 7   paymentMethodProvider           623 non-null    object
 8   transactionAmount               623 non-null    int64
 9   transactionFailed               623 non-null    int64
 10  orderState                      623 non-null    object
dtypes: int64(4), object(7)
memory usage: 53.7+ KB
```

Figure 2. *The structure of the file Customer_DF (1).csv*

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 168 entries, 0 to 167
Data columns (total 10 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Unnamed: 0              168 non-null    int64
 1   customerEmail           168 non-null    object
 2   customerPhone           168 non-null    object
 3   customerDevice          168 non-null    object
 4   customerIPAddress       168 non-null    object
 5   customerBillingAddress  168 non-null    object
 6   No_Transactions         168 non-null    int64
 7   No_Orders               168 non-null    int64
 8   No_Payments             168 non-null    int64
 9   Fraud                   168 non-null    bool
dtypes: bool(1), int64(4), object(5)
memory usage: 12.1+ KB
```

Figure 3. *The structure of the file cust_transaction_details (1).csv*

In a further step of the research activity, we deleted from each file the columns: Unnamed: 0, which contained redundant indexes. Then we merged the files and create a single dataset with the following structure:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 819 entries, 0 to 818
Data columns (total 18 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   customerEmail                      819 non-null    object
 1   transactionId                      819 non-null    object
 2   orderId                            819 non-null    object
 3   paymentMethodId                    819 non-null    object
 4   paymentMethodRegistrationFailure   819 non-null    int64
 5   paymentMethodType                  819 non-null    object
 6   paymentMethodProvider              819 non-null    object
 7   transactionAmount                  819 non-null    int64
 8   transactionFailed                  819 non-null    int64
 9   orderState                         819 non-null    object
 10  customerPhone                      819 non-null    object
 11  customerDevice                     819 non-null    object
 12  customerIPAddress                  819 non-null    object
 13  customerBillingAddress             819 non-null    object
 14  No_Transactions                    819 non-null    int64
 15  No_Orders                          819 non-null    int64
 16  No_Payments                        819 non-null    int64
 17  Fraud                              819 non-null    bool
dtypes: bool(1), int64(6), object(11)
memory usage: 116.0+ KB
```

Figure 4. *The final dataset*

The following steps of the analysis included checking the dataset for missing values. We have found no blank cells in the dataset. The initial analysis of the data produced information that was subject to further in-depth analysis:

```
{'Average number of orders': 3.808302808302808,
 'Average number of payments': 2.1355311355311355,
 'Average number of transactions': 5.1953601953601956,
 'Average transaction value': 35.14774114774115,
 'Maximum number of orders': 8,
 'Maximum number of payments': 15,
 'Maximum number of transactions': 15,
 'Maximum transaction value': 353,
 'Minimum number of orders': 0,
 'Minimum number of payments': 0,
 'Minimum number of transactions': 0,
 'Minimum transaction amount': 10,
 'Number of transactions': 819,
 'Order state unique value': 3,
 'Unique billing addresses': 141,
 'Unique customer devices': 143,
 'Unique customer phone numbers': 143,
 'Unique device addresses': 140,
 'Unique e-mail addresses': 136,
 'Unique payment methods': 4,
 'Unique payment service': 10}
```

Figure 5. *Initial statistics*

The following steps included a more detailed analysis, which showed that the dataset contain 453 suspicious transactions, out of 819. A further step involved checking the data recorded in the various columns against suspicious transactions, including consumer email addresses, telephone numbers, payment methods and payment service providers. While performing the analysis, we developed a chart of transaction amounts concerning payment methods, including credit cards, bitcoin, Apple Pay and PayPal.

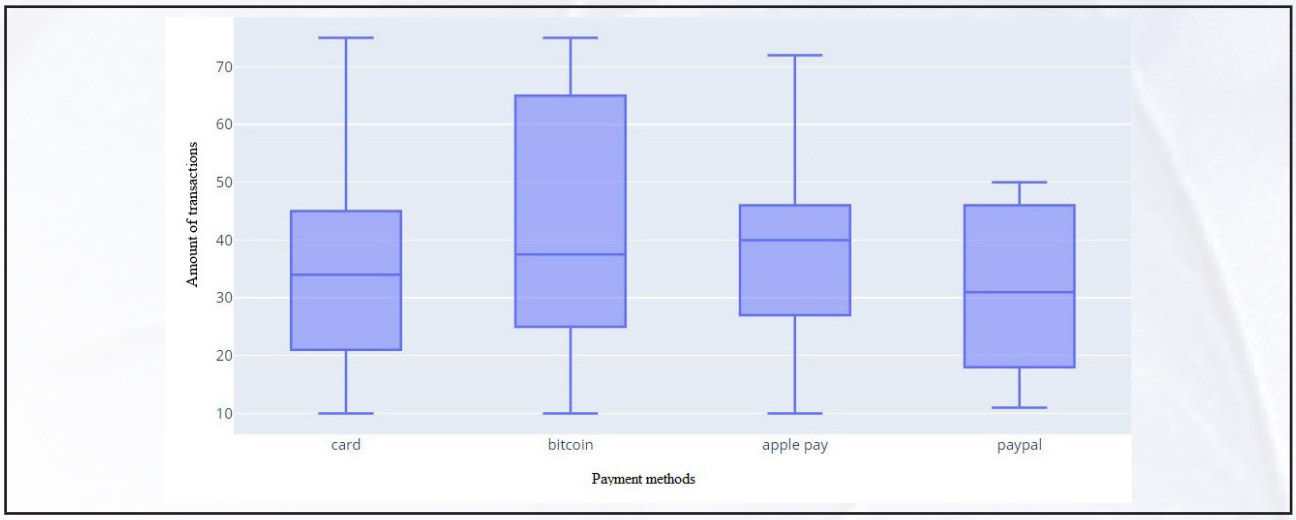


Figure 6. *Amount of transactions in relation to payment methods*

The box plot shows customers made payments in higher amounts with virtual currency. This graph shows that in this case the data are more scattered than in the other cases. On the other hand, customers carried the highest number of transactions with payment cards, as shown by the following figure No. 7. To understand the graph, we marked transactions carried out with payment cards by a circle.

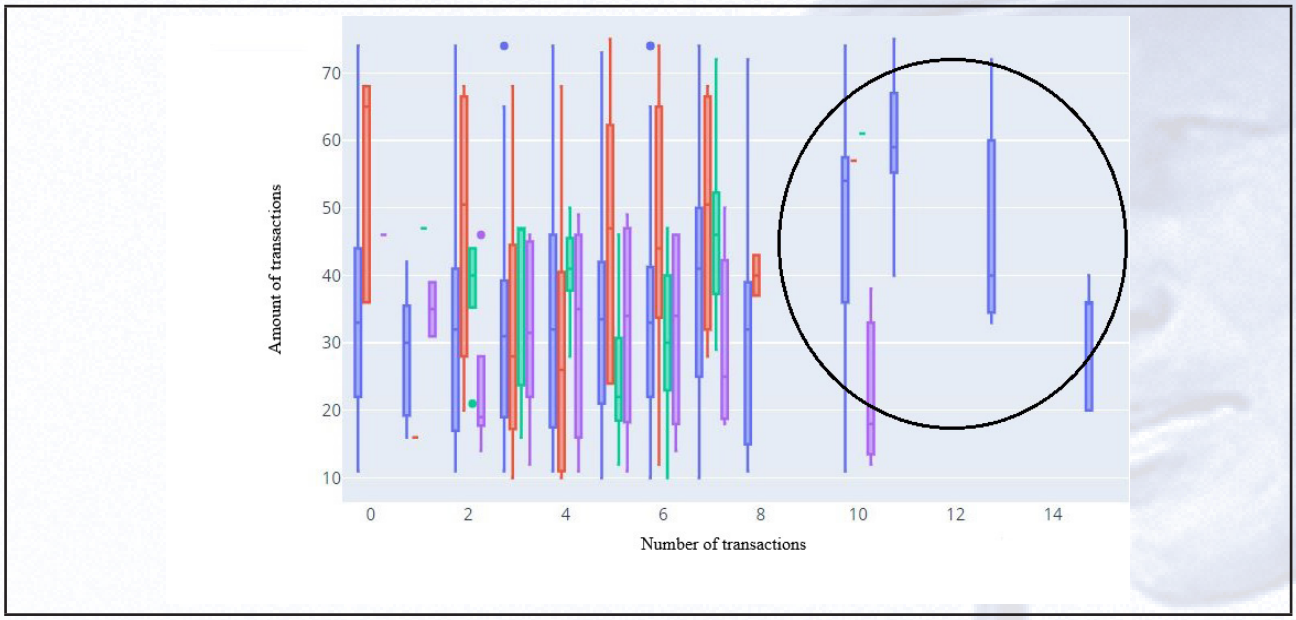


Figure 7. Number of transactions in relations to payment methods.

TRAINING AND EVALUATION OF SELECTED MODELS

To train the prepared models, we defined relevant features with which create dataset: modelling_feature. The structure of the created dataset is following:

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 818 entries, 0 to 818
Data columns (total 8 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   orderState            818 non-null    object
 1   transactionFailed     818 non-null    int64
 2   transactionAmount     818 non-null    int64
 3   No_Transactions       818 non-null    int64
 4   No_Payments           818 non-null    int64
 5   paymentMethodType     818 non-null    object
 6   paymentMethodProvider 818 non-null    object
 7   Fraud                 818 non-null    bool
dtypes: bool(1), int64(4), object(3)
memory usage: 51.9+ KB
```

Figure 8. *The set of modeling features.*

We analysed the set containing the modelling characteristics to identify and remove outliers. Then, we applied coding categories and checked correlations between variables by removing the most correlated features. We further used the scikit-learn library and created a training set and a test set. We standardised the explanatory variables before learning and evaluating the model.

We should emphasise that with the described data set, we are dealing with supervised learning. In supervised learning, the data provided to the algorithm contains the attached solution to the problem, namely the so-called labels. The described in this article contains labels in the Fraud column. The values in this column allow the identification of suspicious transactions. The classical task of supervised learning is the so-called classification. Classification makes it possible to predict class labels in recent occurrences based on past observations. With the financial dataset, these will be new payment transactions.

To solve the problem of classifying electronic payment transactions, we used the following classifiers: SVC, Linear SVC and k-Nearest Neighbours. The accuracy, precision and sensitivity of the first two classifiers was 100%, which was confirmed by model validation. The k-Nearest Neighbours model got the following results:

- accuracy: 0.9329268292682927%,
- precision: 0.9529411764705882%,
- sensitivity: 0.9204545454545454%.

OPTIMISATION

To optimise the k-Nearest Neighbours model, we performed a model parameter search process using the GridSearchCV() method. The got grid of parameters allowed to increase all parameters of the k-Nearest Neighbours model to 100%.

CONCLUSIONS

In about growing globalisation and dissemination of technological solutions, one challenge is development and genuine support of state authorities in realizing effective investigative activities. Identification of perpetrators of economic crimes is painstaking work. This work requires recognition of the mechanism of the crime (modus operandi), disclosure of sources of evidence, and effective collection and proper interpretation of an enormous amount of various data sets. Financial data are among the most important evidential and detection data. It is necessary to provide law enforcement agencies with technological solutions, which will significantly accelerate and improve data analysis. Among the most promising are solutions from the area of artificial intelligence.

This publication is illustrative and shows that machine learning (as an area of artificial intelligence) applies in the automatic in automated analysis of financial data. However, this topic requires further detailed research based on real and good quality financial data sets. Certainly, the quality and quantity of data, alongside other important elements: the number of variables, model parameters and the type of problem, is an element that affects the effectiveness of models used in machine learning.

REFERENCES

1. Criminal Procedure Code of the Republic of Poland, Article 2 (1997). https://www.legislationline.org/download/id/4172/file/Polish%20CPC%201997_am%202003_en.pdf
2. Leehealey, A.& Chirgula A. (2019), Fighting Crime with Artificial Intelligence (AI). In H. R. Arabnia, D. de la Fuente, E. B. Kozerenko & J. A. Olivas, F. G. Tinetti (Eds.) Proceedings of the 2019 International Conference on Artificial Intelligence (pp. 284-290). CSREA Press.
3. Ministry of Internal Affairs and Administration of the Republic of Poland. (2017, October 17). *Report on the state of security in Poland in 2016.* https://archiwumbip.mswia.gov.pl/bip/raport-o-stanie-bezpie/18405,Raport-o-stanie-bezpieczenstwa.html
4. Nepelski, M., & Struniawski, J. (Eds), *Profilaktyka bezpieczeństwa*, Wydawnictwo Wyższej Szkoły Policji w Szczytnie.
5. Polish National Police. (n.d). Statistics. *Economic Crimes.* Retrieved July 2, 2021, from *https://statystyka.policja.pl/st/przestepstwa-ogolem/przestepstwa-gospodarcz/122291,Przestepstwa-gospodarcze.html*
6. Rastogi A. (2020), *Ecommerce Fraud Data*. Retrieved July 1, 2021, from https://www.kaggle.com/aryanrastogi7767/ecommerce-fraud-data/metadata.