

Automatic Detection of County Lines Criminal Scheme

Tomasz Granosik* and Elżbieta Lewańska

Department of Information Systems, Institute of Informatics and Quantitative
Economics, Poznan University of Economics and Business,
al. Niepodległości 10, 61-875 Poznań, Poland

Tomasz.Granosik@ue.poznan.pl, Elzbieta.Lewanska@ue.poznan.pl

Abstract. County lines are a new type of criminal activity in the UK in which young people are being forced to participate in drug dealing. This article presents an application of an unsupervised machine learning technique to detect such cases among a bank's clients, using financial and spatial data. It proposes and presents a system for detecting county lines crime schemes and uses the example of integrating spatial analysis into financial crime detection. The initial results will be a base for further research.

Keywords: County Lines, Crime, Money Laundering, Banking, GIS, Spatial Analysis, Anomaly Detection.

1 Introduction

Most of our everyday activities leave a geographic trace and financial behavior is no different. Information about geographic location can be a valuable addition to existing financial crime detection techniques. However, to the best of the authors' knowledge (as outlined in Section 2), detailed spatial analysis has rarely been used in the banking industry and in machine learning models in particular.

Making sudden, unusual transactions in places away from the country of residence can raise alarms about a client in financial institutions. However, it seems that current solutions are based on business defined rules on the national, or even broader, level. This distinguishes the worlds of data science and Geographic Information Systems (GIS) despite their having a number of similarities. This work tries to narrow that gap whilst, at the same time, proposing a system for detecting county lines – a new but threatening type of crime emerging in the United Kingdom [1], [2].

Traditionally, low level drug dealing gangs (i.e. not involved in production and cross border trafficking, but distribution to end users) have operated on a limited, local scale. Nowadays they have started to enhance their area of operations from big cities into the countryside, toward suburbs and smaller towns. These new areas usually have good transport links to major urban centers and include neighborhoods in which neither deprivation nor affluence are notably prevalent [2].

* Corresponding author

This type of crime typically involves the supply of heroin and crack cocaine although other types of substances are also encountered [3]. The orders are received using either a single or multiple dedicated phone lines controlled by senior criminals. These numbers can be advertised within the target groups and act as a brand.

One of the key features of this scheme is that organized gangs are using young or vulnerable (mainly due to health difficulties) people as drug traffickers and money mules, transporting substances and proceeds between various points and stages of the operation [2].

The demographic groups at risk are predominantly males aged between 15 and 19 [1] with known cases ranging from 11 to 56 years old and including both sexes. Individuals used in such a fashion may be convinced by an offer of financial gains, luxury items or free drugs which may be immediately, or later, followed by threats, violence or debt bondage. As such, it is closely connected to sexual exploitation and modern slavery. Another activity related to county lines is cuckooing. Apart from using their victims as runners, gang members may move into their houses and set up the base of operation there [4].

Whilst these kinds of people are an essential part of the crime and are involved in committing felonies, they are not the ultimate financial beneficiaries (if they receive any gains at all) and are treated as victims in this article, whereas the term criminals is reserved for their overseers and kingpins.

The individuals involved will use various modes of transport – most often trains (around 40% of the time [3]), with a significant number using coaches, cars, taxis and similar services. In order to avoid suspicion one journey may involve several modes of commuting. The largest number of lines originate in the biggest agglomerations: London, Manchester, Birmingham and Liverpool.

It is estimated that, in 2019, 2000 phone numbers were being used in drug trafficking in the UK [5] and that the yearly profit from operating just one line can exceed 800 000 GBP, with the involvement of 27 000 young people in total [1].

Whilst the crime described above has multiple consequences, affecting various members of society, its detection faces several obstacles:

- The crime itself is fairly new and less researched than older, more traditional types of felonious patterns.
- Victims (people used to move drugs and money) and clients may be reluctant to cooperate with enforcement agencies due to the illegal activities that they are involved in and the fear of being held responsible.
- Large area of operations may include multiple police areas which require coordination on a higher organizational level.
- Criminals are enhancing their operations into rural and coastal areas which have limited enforcement resources.
- Whilst police and other government agencies may have access to criminal records, they can obtain access to financial data only after following a legal procedure.

Banks and other financial institutions are taking measures to detect financial crimes, especially money laundering, and the use of illegally obtained funds, in line with external and internal policies. They already gather and store information on financial activities of their customers, which they can process in compliance with the law, including data protection regulations.

It is important, then, from legal and ethical perspectives, to develop a system that will be able to identify malicious activity on customers' accounts. This can be enhanced by a novel approach – using real data, a large set of spatial points on clients' locations and the actual business scenario. Combining Geographical Information Systems (GIS) in the financial industry has not previously been adequately researched, as described in Section 2.

This work is a part of ongoing research. The verification of the results by experts has not been finished at the time of writing. As such, the article will focus on presenting the problem, proposed methods and preliminary results.

The aim of this article is to present an example of aligning the financial industry with IT by supporting the work of the Financial Crime Investigation Unit in a bank by means of a machine learning solution. In particular, the goals are twofold:

- To propose a system for detecting county lines crime schemes.
- To present an example of integrating spatial analysis into financial crime detection.

The remaining parts of this article will be structured as follows: the next section will present a survey aimed at identifying the state-of-the-art integration of geographic features into financial crime detection and data science. Sections 3, 4 and 5 will present the modelling approach, the data, and the application of spatial information, respectively. Section 6 will outline the results and their evaluation, while Section 7 will discuss their applicability and the next steps.

2 Related Work

The review below aims to answer two research questions, i.e.: How is GIS used in the financial industry and how are spatial analyzes used to detect and combat financial crimes?

Four digital libraries have been used in this survey:

1. arXiv[†]
2. Science Direct[‡]
3. Scopus[§]
4. Springer Link^{**}

They have been chosen based on their availability, search and filter options, ease-of-use and contents.

The list of keywords has been specified in Table 1. The search terms have been combined using Boolean operators and parentheses to produce more complex queries, e.g. fraud AND (“systematic literature review” OR SLR OR “comprehensive literature review”).

Table 1. The keywords used during the review

Area	Keywords
GIS	GIS, geographic information
Financial sector	Banking, financial crime, fraud, card, insurance, money laundering, finance
Other	Anomaly, outlier, detection, SLR, systematic literature review, comprehensive literature review, machine learning

After conducting the keywords-based search, the results have been filtered using the following inclusion criteria:

- Articles have been published in English.
- Articles have been published in 2011 or later. This condition has been chosen to ensure that the information represents the state of the art as 2011 was the year when Hadoop was released and thus it can be interpreted as a symbolic beginning of the big data era. However, some of the previously known texts about geospatial analyzes were included even if they were published earlier. The rationale for that decision is that the risk of excluding works on a topic outweighs the consequences of inclusion of an older, possibly outdated, paper. An additional remark would be that if the older works have not been followed up and repeated in more recent years, then this further strengthens the assumption that the area in question is poorly researched.

[†] <https://arxiv.org/>

[‡] <https://www.sciencedirect.com>

[§] <https://www.scopus.com>

^{**} <https://link.springer.com>

- Document type is either a peer reviewed article, a conference proceeding or a book chapter, if it had the features of a scientific article. Again – this was not strictly enforced in the case of queries about GIS, due to the lack of results.
- The scientific area has been defined by authors or publishers as computer science, economics, business management, law or crime. In contrast – life sciences, ecology, urban planning etc. have been used as an excluding criterion, – as GIS is a frequently used tool in these areas, omitting them helped to reduce the number of false positives.

Following that step, articles have been assessed based on the relevancy of their title, keywords, abstract and conclusions (in this sequence). The accuracy of the evaluation heavily depended on the journal and publisher, as sometimes keywords have been created automatically whereas abstracts did not always provide sufficient overview. It has been assumed, however, that all of these features together have given a sufficient overview.

The retained, distinct entries have been read and assessed based on their full text.

As papers about combining GIS and financial analyzes have been found to be rare, the results of the systematic review have been enhanced by previously known works. To ensure appropriate diligence, general search engines have been used to find materials which had at least part of the properties of scientific articles (such as approach or detailed descriptions), but were published in non-peer reviewed press, or even commercial, grey literature. It has been considered and decided that the risk of including works that may not adhere to scientific principles is outweighed by the additional knowledge brought by them. The other conditions have likewise not been strictly applied, meaning that works published before 2011 could also be included if they presented a novel approach.

The five points specified below were used to assess the articles during their final evaluation, although not all of them had to be met in order to include the paper in the review.

- The use of real-world datasets has been considered an important feature that indicates the practical application of the proposed solution: this is because artificially added examples of fraud have to be considered an oversimplification.
- Has the work included geospatial analyzes?
- Has the work described financial crime detection?
- Have the problems and solutions been described in a clear and repeatable manner?
- Has the paper presented clear and practical conclusions?

The results are presented below.

One of the most obvious applications of GIS is finding the optimal location of various facilities using multiple criteria. For banks that usually means identifying the sites for new ATMs or branches and is described quite well in the literature [6]–[10]. Using multiple layers with spatially located data such as population density, socioeconomic factors (e.g. education and employment), social, commercial and transportation facilities – schools, shopping centers, roads, land prices etc., it is possible to narrow down the available choices to just a few locations. Such a method can also be used to check, not only the best location for new branches, but also the lack of existing ones [11], which, in turn, can lead to the identification of vulnerable communities and neighborhoods at risk of social exclusion.

One example of graph based solutions to money laundering [12] has calculated geographical risk to be at the same rather high level as Italian regions and foreign states. This has taken into account variables such as crime rate, number of suspicious operations, organized crime activity, index of corruption and listed tax havens.

Ardizzi et al. [13] have proposed a solution for detecting money laundering and analyzing cash anomalies in Italy with the degree of resolution set to second and third level administrative units. The use of spatially located data as input for the model has been limited to dummy variables. The aggregated results have been visualized on maps.

Geography has been identified as an important demographic variable in an Australian approach [14]. The scope of the analysis was the detection of demographic patterns of money

mules – one of the stages that money launderers use to hide the real source of their funds. The level has been quite detailed as it was set to Australian postcodes (of varying size but with an average population of ten thousand people). The data suggested that the largest urban centers were more vulnerable to money laundering than would be suggested by simply looking at population density. Additionally, the study has identified social factors that increased the risk of involvement in illegal financial activities.

Another example of using geographical information systems to combat money laundering has been described in an ESRI whitepaper [15]. It mentioned that low value money orders can be used to transfer financial assets anonymously and to hide the illegal source of funds (in a practice known as smurfing). In order to combat this, the United States Postal Service (USPS) has been trying to detect where anomalies and suspicious behaviors were taking place and to link the parties who were involved. While there has been no mention of machine learning solutions, it did indicate the potential for using maps in the financial world. Visualizing the hotspots and transfer patterns can help Anti-Money Laundering (AML) officers to identify places of interest, e.g. postal offices where exceptionally high numbers of money orders were bought.

In a similar approach, geoinformatics has been applied to indicate the possible areas at increased risk of money laundering [16] in the Netherlands. Maps have been used to show the outcomes of auditing the enforcement of AML policies at the level of Dutch Security Regions (25) in the whole country. The results indicated several inadequacies and differences between provinces in numbers of financial crimes, as well as their detection and punishment rates.

A very good example of the potential in applying GIS to challenges faced by the financial world has been presented in an Association of Certified Anti-Money Laundering Specialists (ACAMS) whitepaper [17]. This has combined financial data, such as the estimation of funds involved in money laundering, with typical GIS features – population density, rail networks, urban agglomerations, border crossings etc.

One of the case studies presented an approach for identifying high risk customers that might be involved in corruption and terrorism financing. They were defined as Turkish border guards and customs officials working at the crossings opposite the Syrian areas occupied by the so-called Islamic State. Another case showed the method for identifying bank branches at risk of being exposed to money laundering attempts, due to their proximity to areas of terrorism activity, such as northern Nigeria and Somalia.

GIS has also been used in areas such as assessing the performance of financial institutions and the state of the economy. Alem & Townsend [18] showed a positive impact of access to banks and their products on the financial state of Thai households. In the survey, geospatially located data have been used as one of the factors in the regression-based model used as a measure of access to financial institutions at the level of single villages.

Spatial analyzes have also been proven to work in the research of socio-economic and demographic factors. In a study conducted for Milan [19], mobile phone activity has been found to have a high correlation with the presence of various points of interest (including the temporal ones such as sporting events) and the patterns of usage of the international calling codes showed the distribution of various ethnicities within the city. The data was based on regular cells, each of 400 m², covering the urban area under study.

Another proven area of application is the detection of mortgage frauds and prediction of housing values. Whilst this was one of the causes of the financial crisis in 2008, a similar approach worked also for historical data [20]. The study conducted [21] showed geographic patterns of high value mortgages on low price real estate in Charlotte-Mecklenburg County in the US, indicating the possibility of a fraudulent scheme. The scope of the analysis included calculating the distance between the parcels of land, but also aggregating them to the level of neighborhoods (acting as statistical division units).

The places where a customer has used a card, together with their order and frequency, have been used as important inputs to card fraud detection [22]. The authors have used an association rules algorithm to check patterns of behavior and created a directed graph based on a conditional

probability that the next transaction would happen in a given location. The approach has not used any spatial analysis (such as distance or direction between locations) in a narrow sense and, as noted, detects deviations only from very regular patterns, which may not necessarily be fraudulent.

The initial results of this survey indicate that there might be an area for improvement and research. Applying GIS to financial crime prevention has the potential to be a novel approach with a major role in real world applications. The important factors should be the practical application of the proposed solution using the actual data which banks are justified in gathering and for which they have processing capabilities.

The scale of analysis varied from the national level to single households. While the former is too general, the latter is certainly the most detailed, but it is not all data that is gathered and available at this degree of detail. The answer to this challenge may be the use of small administrative, statistical or postal units encompassing a few blocks within a city or a comparable area.

3 Anomaly Detection Method

As mentioned in Section 1 county lines crime has not been adequately researched. This means that, at the time of conducting the analysis, there were not enough identified cases to create reliable data labels. Because of this an unsupervised algorithm has been used.

The assumption in the proposed approach was that the default behavior of customers, common to the vast majority of them, is legal and unsuspecting. As a consequence, the data reflects their typical, lawful, everyday activities, spending and locations visited on a normal basis without criminal intent. This forms the set of inliers. A significant deviation from this pattern would indicate an anomaly, a behaviour that is out of the norm.

The features used in this application were selected, based on the knowledge of experts, as possible indicators of county lines related activities. As such, the identified outliers could be treated as cases with a high risk of involvement in drug dealing.

The exact choice of the algorithms depended on the properties of the dataset – sparse, with multiple dimensions (44 columns) having varying distributions. Additionally, as the prediction may have legal consequences (i.e. escalating a customer's case to the police), it has to be explainable. For that reason, for the purpose of this analysis, the isolation forest [26] algorithm, as implemented in scikit-learn Python package [24], has been chosen.

This is an ensemble-based method that can be treated as an unsupervised version of random forest. It identifies anomalies as points that can be described using fewer rules than inliers.

One of the main parameters of the model is that contamination should be set to the proportion of the expected anomalies. Because the estimation of the number of cases in a population of customers was not available, another method had to be used for the initial value. The assumption was that the share of people involved in county lines is the same amongst the bank's clients as in the general population. It is estimated that up to 27 thousand youngsters may be part of drug dealing gangs in England and Wales [1] out of a total population of 3.5 million people aged 15 to 19 years old [27], so the parameter's initial value has been set to 0.007.

In order to ensure that no single variable will have an overwhelming impact on the model's output, one, randomly chosen, feature has been excluded from training each estimator tree. This has been governed by the `max_features` parameter.

Taking the above into consideration, the model has been trained and scored on the whole dataset. As a result, each data point has been assigned a label: -1 for outliers and 1 for inliers, as well as a continuous score with anomalies marked with a negative value.

It is important to underline that any cases detected as anomalies should not be automatically labeled as criminal behavior. They deviated from the majority and shared some patterns with the expected features of county lines activities – e.g. high daily turnover, cash withdrawals from multiple locations and large spending on transportation. However, even if suspicious and

unusual, they might be completely legal and therefore, the final assessment should be made by subject matter experts, following established procedures and ensuring that the processes follow both government and internal policies.

4 Data

The research has used a real set of customers from one of the biggest British banks and a history of one year's worth of their transactions. Only people between the ages of 13 and 20 or those who willingly disclosed their health disadvantages, have been within the scope designated as a group at risk of being involved in county lines activities [1] – over a million of them in total.

It is important to add, from a privacy and data protection point of view, that this approach requires neither the use of any new sources, nor the gathering of any additional information on clients, nor tracking their location, e.g. by a mobile application. This analysis operates only on existing data from core banking processes.

Based on the description of the county lines activities and experts' knowledge, several financial features have been identified as payments for specific types of services, such as mobile top-ups. For each category a series of statistics (sum, maximum, variance etc.) have been calculated. This has been enhanced by indicators of financial behavior inconsistent with age, created following a discussion with subject matter practitioners.

In addition, that information has been linked with a general overview of customers' credits and debits summaries as well as their tenure as clients. Data has been aggregated on a monthly level for each customer.

Before feeding the set into the model it has been standardized to a normal $\sim N(0,1)$ distribution. This pre-processing method has a positive effect on outlier detection techniques, whilst being sensitive to anomalies – something which has been considered an advantage in this particular application [23], [24].

5 Spatial Information

The spatial features are based on locations of three types of events:

- The customers' visits to bank branches.
- Money withdrawals from cash machines.
- The transfers of funds ordered at Post Office branches.

The first two types of events were combined as places physically visited by the customers who were analyzed.

The Post Office transfers were treated separately, as they show the locations from which the funds have been received. Such actions could be made by parties other than the accounts' owners.

The respective spatial features have been analyzed only if a customer visited or received funds from at least three different sites in a given month.

The locations of all of these points have been set to the centers of their postcodes. The system of such areas used in the UK is quite detailed. There are over 1.7 million postcodes currently used in total and, on average, each one encompasses only 36 people and 0.14 km². This has been considered as the right scale giving the sufficient level of details, whilst, at the same time, allowing for more generalized results.

Based on the coordinates, the following two new input features have been calculated per each location type:

The spread of points has been measured using Standard Distance (1) where x_i and y_i are coordinates (northing and easting) of the i -th point, \bar{X} and \bar{Y} are mean values of the coordinates and n is the number of points. This measure can be interpreted as a spatial equivalent of standard deviation [25].

$$SD = \sqrt{\left(\frac{\sum(x_i - \bar{X})^2}{n}\right) + \left(\frac{\sum(y_i - \bar{Y})^2}{n}\right)} \quad (1)$$

The spread of points gave information about the area and variability of a customer’s financial activity. However, it didn’t describe the shape of it. For that reason a new metric has been created that can distinguish the customers who travel between set destinations (e.g. daily commute from home to school or work) and those whose journeys were more diverse. This has been achieved by fitting a line to the points, using the least squares method, and then calculating the absolute value of the coefficient of determination (R^2) as a measure of linearity. Values close to 1 suggested that the points of interest lay along one route, while results closer to 0 indicated travel to multiple places, which is more consistent with the expected behavior of county lines runners.

6 Results and Discussion

The results have been shown in Figure 1. For the purpose of visualizing them in this article, the number of dimensions has been reduced to two, using Principal Component Analysis (PCA). The figure shows a random sample ($n=10\,000$) of the results with 79 outliers; the vast majority of observations, predicted as inliers, are condensed in a small area of the plot.

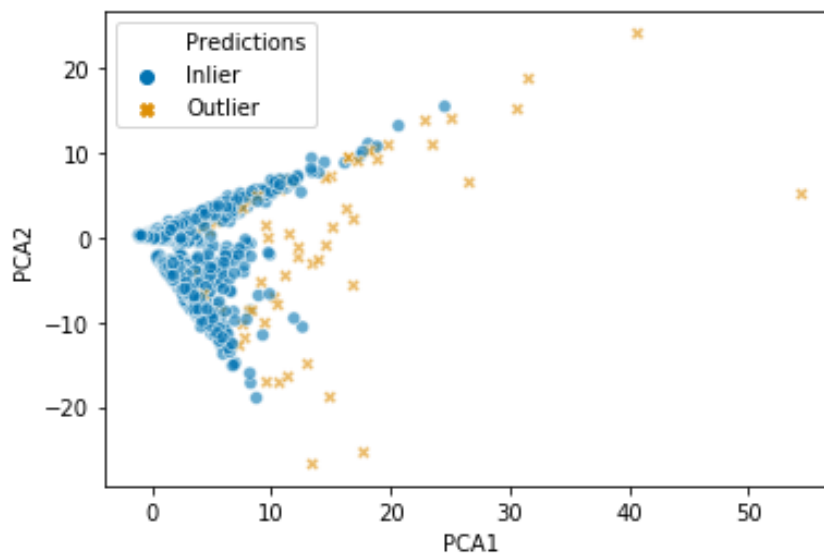


Figure 1. The results of the anomaly detection. PCA1 and PCA2 are two dimensions created using Principal Component Analysis

Anomaly Scores have been plotted in Figure 2. Their values have been based on a mean number of data splits required to identify a particular observation as described in Liu et al. [26] and implemented by Pedregosa et al. [24]. Negative values represent the 0.7% of observations predicted to be anomalies. Scores above 0 indicate inliers. The bottom plot shows the data on a logarithmic scale.

An example of two cases – one identified as an anomaly, another one as an inlier – has been presented in Figure 3. The locations of the outlying case are scattered over a much larger area, inconsistent with the daily commute patterns.

As mentioned before, the threshold (contamination parameter) for anomaly detection was initially set to 0.007 (based on statistical data on county lines frequency in the whole UK). However, it might not be valid for the population of the bank’s clients. The threshold should be adjusted according to the analysts’ information needs – higher values will detect more cases as anomalies but might increase the false positive rate. On the other hand, lower values will lead to

classification of some anomalous behavior as normal. As the proposed method goal is not to bring charges directly, but rather serve as an early-warning system for potential frauds, the final user might decide to use a threshold value that is adjusted to risk appetite. Calculation of the optimal threshold value, depending on different application scenarios, is planned as part of future work.

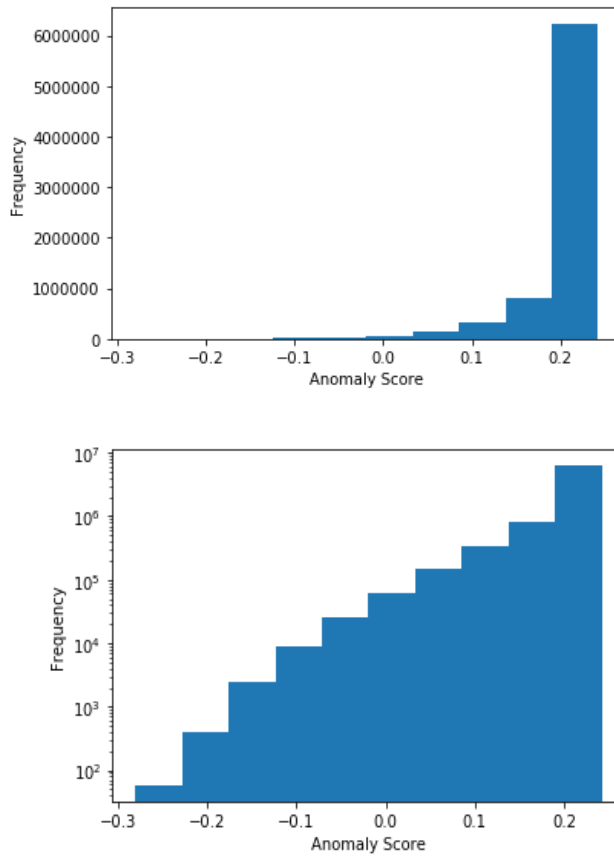


Figure 2. Histograms of Anomaly Scores



Figure 3. The comparison of locations visited by clients predicted by the described model as anomalous (left) and normal behaviour (right). The scales of maps A and B are different.

Since there were no labeled data available to calculate the precision and recall of the results, it is planned to interview a group of subject matter experts. At the time of writing this article this approach has not been completed. However, it will be done as soon as such an option will

become feasible. This method will enable the validation of the detected anomalies, as well as assess the usability of different features, parameters (e.g. contamination) and algorithms. Once the appropriate number of labeled cases is available, the problem can be converted to supervised learning.

Although the verification, in terms of comparison with ground truth, was not possible, the results could be evaluated based on internal patterns within the dataset. Whilst the outliers formed a diverse and scattered set, the inliers were expected to have a high cohesion. This has been checked using Silhouette Coefficient which measures whether the samples in the same group are more similar to each other than to the other set. Its values are bounded by -1 and 1, with the latter indicating the best separation. The coefficient has been calculated for observations predicted to be inliers from a sample of 10000 observations. The mean value of Silhouette Coefficient was 0.94 with a median of 0.96. As both values are close to 1, it shows that the algorithm could differentiate normal and anomalous cases, as expected. It also supports the initial assumption that the group of inliers, prevailing in numbers, forms a condensed and separate set.

7 Conclusions and Future Work

While the labeled data was not available, the next steps will include evaluating the results of the unsupervised algorithms using experts' assessments. An adequate number of confirmed cases will enable switching to a supervised method, aimed specifically at the county lines type of crime.

However, even in the form of anomaly detection, enhanced with the appropriate visualisation of the results, the method presented can be used as a support for financial crime investigators. This will lead to a better detection rate and, in turn, enhanced financial safety for clients and general security for society as a whole.

This paper has also presented an example of advantages resulting from using detailed geographic locations as data sources for machine learning. The data about the locations which customers visited, or from which they have received money, has been converted into numerical features that could be aggregated, compared and used in a model. While being only a small step forward, it creates an important argument for broader integration of GIS and data science.

References

- [1] C. Black, "Review of Drugs Executive Summary," Contains public sector information licensed under the Open Government Licence v3.0., 2020.
- [2] National Crime Agency, "NCA Intelligence Assessment: County Lines, Gangs, and Safeguarding," 2015.
- [3] National Crime Agency, "County lines drug supply, vulnerability and harm 2018 (Intelligence Assessment NAC (19)095)," 2019.
- [4] J. Spicer, L. Moyle, and R. Coomber, "The variable and evolving nature of 'cuckooing' as a form of criminal exploitation in street level drug markets," *Trends Organ. Crime*, Springer, 2019. Available: <https://doi.org/10.1007/s12117-019-09368-5>
- [5] Ministry of Justice, "County Lines Exploitation Practice guidance for YOTs and frontline practitioners," October, 2019.
- [6] S. Allahi, M. Mobin, A. Vafadarnikjoo, and C. Salmon, "An integrated AHP-GIS-MCLP method to locate bank branches," *IIE Annu. Conf. Expo 2015*, pp. 1104–1113, 2015.
- [7] A. Basar, Ö. Kabak, Y. Topçu, and B. Bozkaya, "Location Analysis in Banking: A New Methodology and Application For a Turkish Bank," *Applications of Location Analysis. International Series in Operations Research & Management Science*, vol. 232, Springer, pp. 25–54, 2015. Available: https://doi.org/10.1007/978-3-319-20282-2_2
- [8] J. G. Cabello, "A decision model for bank branch site selection: Define branch success and do not deviate," *Socioecon. Plann. Sci.*, vol. 68, p. 100599, 2019. Available: <https://doi.org/10.1016/j.seps.2017.09.004>

- [9] A. Mimis, “A geographical information system approach for evaluating the optimum location of point-like facilities in a hierarchical network,” *Geo-Spatial Inf. Sci.*, vol. 15, no. 1, pp. 37–42, 2012. Available: <https://doi.org/10.1080/10095020.2012.708156>
- [10] N. A. Pugi, H. Zainol, and A. A. Mohamad, “Determination of New Bank Branch Location Using GIS Approach,” *Regional Conference on Science, Technology and Social Sciences (RCSTSS 2014)*, Springer, p. 201–208, 2016. Available: https://doi.org/10.1007/978-981-10-0534-3_19
- [11] S. W. Hegerty, “Commercial bank locations and ‘banking deserts’: a statistical analysis of Milwaukee and Buffalo,” *Ann. Reg. Sci.*, vol. 56, no. 1, pp. 253–271, 2016. Available: <https://doi.org/10.1007/s00168-015-0736-3>
- [12] A. F. Colladon and E. Remondi, “Using Social Network Analysis to Prevent Money Laundering,” *Expert Syst. Appl.*, vol. 67, pp. 49–58, 2017. Available: <https://doi.org/10.1016/j.eswa.2016.09.029>
- [13] G. Ardizzi, P. De Franceschis, and M. Giammatteo, “Cash payment anomalies and money laundering: An econometric analysis of Italian municipalities,” *Int. Rev. Law Econ.*, vol. 56, pp. 105–121, 2018. Available: <https://doi.org/10.1016/j.irle.2018.08.001>
- [14] M. Aston, S. McCombie, B. Reardon, and P. Watters, “A preliminary profiling of internet money mules: An Australian perspective,” *UIC-ATC 2009 – Symp. Work. Ubiquitous, Auton. Trust. Comput. Conjunction with UIC’09 ATC’09 Conf.*, pp. 482–487, 2009. Available: <https://doi.org/10.1109/UIC-ATC.2009.63>
- [15] ESRI, ArcNews, esri.com, 2008.
- [16] E. Jongsma and F. De Graaf, “Audit Efforts to Combat Money Laundering and Terrorism Financing,” *INTOSAI – International Journal of Government Auditing*, vol. 35, no. 3, pp. 6–12, 2008.
- [17] M. Wakeman, “Exploring Potential Uses of Geographic Information Systems and Predictive Analysis in AML/CTF Investigations,” 2015.
- [18] M. Alem and R. M. Townsend, “An evaluation of financial institutions: Impact on consumption and investment using panel data and the theory of risk-bearing,” *J. Econom.*, vol. 183, no. 1, pp. 91–103, 2014. Available: <https://doi.org/10.1016/j.jeconom.2014.06.011>
- [19] P. Bajardi, M. Delfino, A. Panisson, G. Petri, and M. Tizzoni, “Unveiling patterns of international communities in a global city using mobile phone data,” *EPJ Data Sci.*, vol. 4, no. 1, pp. 1–17, 2015. Available: <https://doi.org/10.1140/epjds/s13688-015-0041-5>
- [20] G. Meen, K. Gibb, C. Leishman, and C. Nygaard, *Housing economics: A historical approach*, Springer, 2016. Available: <https://doi.org/10.1057/978-1-137-47271-7>
- [21] R. E. Wilson and A. D. Fulmer, “Using Near Repeat Analysis for Investigating Mortgage Fraud and Predatory Lending,” *Forensic GIS. Geotechnologies and the Environment*, vol. 11, Springer, pp. 73–103, 2014. Available: https://doi.org/10.1007/978-94-017-8757-4_5
- [22] C. C. Lee and J. W. Yoon, “A data mining approach using transaction patterns for card fraud detection,” 2013.
- [23] S. Kandanaarachchi, M. A. Muñoz, R. J. Hyndman, and K. Smith-Miles, “On normalization and algorithm selection for unsupervised outlier detection On normalization and algorithm selection for unsupervised outlier detection,” *Data Min Knowl Disc.*, vol. 34, Springer, pp.309–354, 2019. Available: <https://doi.org/10.1007/s10618-019-00661-z>
- [24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [25] A. Mitchell, *The ESRI Guide to GIS Analysis*, Volume 2. ESRI Press, 2005.
- [26] F. T. Liu, K. M. Ting, and Z. H. Zhou, “Isolation forest,” *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 413–422, 2008. Available: <https://doi.org/10.1109/ICDM.2008.17>
- [27] Office for National Statistics, “2011 Census aggregate data. UK Data Service (Edition: June 2016),” 2016. Available: <https://doi.org/10.5257/census/aggregate-2011-2>