Prediction of Crime Occurrence in case of Scarcity of Labeled Data
Etiketlenmiş Verilerin Kıtlığı Durumunda Suç Oluşumunun Tahmini

Volkan Kiranoglu [1] , Goksu Tuysuzoglu [1*] , Elife Ozturk Kiyak [1]

[1] Dokuz Eylul University Department of Computer Engineering, Izmir, TURKEY
Corresponding Author *: goksu.tuysuzoglu@ceng.deu.edu.tr

Abstract

In line with technological developments, machine learning/data mining studies have significantly scaled up in crime analysis. The prediction of crime occurrences, the detection of the spatial/temporal distribution of the criminal cases, forecasting the type of crime are some of these study areas. By taking crime data resulting from a substantial increase in crime rates into consideration, unlabeled data can be utilized to enhance exploring the patterns of crime for future events or to make crime-related predictions easily. Therefore, in this study, active learning, self-learning, and random sampling techniques are applied to predict the outcome of criminal searches in England using the police data of 2019. According to the experimental analysis, active learning outperforms its counterparts using its entropy-based smart selection strategy data in case there is little labeled data.

Keywords: Active Learning, Classification, Crime Detection, Random Sampling, Semi-Supervised Learning, Self-Learning

Öz

Teknolojik gelişmeler doğrultusunda, makine öğrenmesi/veri madenciliği çalışmaları suç analizinde önemli ölçüde artmıştır. Suç olaylarının tahmini, ceza davalarının mekansal/zamansal dağılımının tespiti, suç türünün öngörülmesi bu çalışma alanlarından bazılarıdır. Suç oranlarındaki önemli artıştan kaynaklanan suç verileri dikkate alındığında, gelecekteki olaylar için suç kalıplarını araştırmak veya suçla ilgili tahminleri kolayca yapmak için etiketlenmemiş veriler kullanılabilir. Bu nedenle, bu çalışmada, 2019 polis verilerini kullanarak İngiltere'de suç araştırmalarının sonucunu tahmin etmek için aktif öğrenme, kendi kendine öğrenme ve rastgele örnekleme teknikleri uygulanmıştır. Deneysel analize göre, aktif öğrenme, çok az etiketlenmiş veri olması durumunda düzensizliğe dayalı akıllı seçim stratejisini kullanarak muadillerinden daha iyi performans göstermektedir.

Anahtar Kelimeler: Aktif Öğrenme, Sınıflandırma, Suç Tespiti, Rastgele Örnekleme, Yarı Denetimli Öğrenme, Kendi Kendine Öğrenme

1. Introduction

Crime analysis, which handles crime, suspect, and victim in a systematic way, is really important to take the needed precautions and to determine the appropriate protection techniques accordingly because crime cases highly influence the reputation of a region,

social lives of the habitants and the economic growth of a country. Since the subject is significant, many machine learning and data mining studies are also carried out in this field. Prediction of crime cases [1-3], estimation of the crime type based on crime features [4], prediction of the identity of the suspect [5-6], forecasting the spatial distribution of crime events [7] are some of these study areas, where basic tasks of data mining, i.e. classification, clustering, regression, and association rule mining techniques, were applied.

In this direction, the aim of this study is to predict the outcome of criminal events (whether the accused is guilty or not) using the past police data of England recorded in 2019 by applying active learning, self-learning, and random sampling techniques. Four well-known classifiers, decision tree (DT), naïve Bayes (NB), support vector machines (SVM), and k-nearest neighbors (KNN), are applied as base learners of the constructed models. Although semi-supervised learning methods have been applied to many areas, to the best of our knowledge, there are no studies that compare self-learning, active-learning, and random sampling.

The main contributions of this paper are listed as follows:

- The applied techniques are compared to each other so that which method is more suitable for this type of study is tested.
- Accurately identifying new crime cases becomes easier by capturing similar patterns using machine learning approaches.
- This study makes possible that in case there are criminal cases, where the outcome of few is detected, the pool of the labeled cases (i.e. the outcome is true or false) can be expanded using the hidden patterns in the pre-labeled ones by applying a semi-supervised learning technique to the unlabeled cases.
- This study can offer an auxiliary support for the authorities in police departments to estimate the results of criminal events more easily.

The rest of this paper is designed as follows. Section 2 addresses the recent data mining/machine learning studies presented in crime detection and the application areas of the semi-supervised learning and active learning techniques used in this study. In the next part (Section 3), brief explanations of the applied methodologies (active learning, self-learning, and random sampling) are given. The data set of experimental studies is introduced in Section 4; furthermore, the results of data analysis showing some statistics are shown. Experimental results of machine learning algorithms are explained in Section 5. In the part of the Conclusion, a general summary of the study and future work are given.

2. Related Work

This section explains the recently proposed studies in two-folds: machine learning/data mining studies proposed for crime analysis and the application areas of semi-supervised learning and active learning techniques.

2.1. Recent studies on crime analysis

Crime analysis, which aims to reduce the crime rate, involves predicting crimes and exploring patterns of crimes. Therefore, studies in the literature have focused on these two objectives applying various machine learning algorithms.

To predict crime categories or patterns, classification is the most commonly used technique. Iqbal et al. [8] utilized Decision Tree and Naïve Bayes algorithms to predict the crime category in different regions. To predict crime type in a specific area, Nguyen et al. [9] used Support Vector Machine, Random Forest, Gradient Boosting Machines, and Neural Network. In addition to classification, regression was also useful in crime prediction. Alves et al. [10] applied a random forest regressor to predict crime and see the impact of urban information on murders.

In addition to classification and regression, to explore crime patterns, Chhabra et al. [11] have proposed an approach which is a combination of hierarchical and k-means clustering algorithms. And also, in [12] C-Means clustering algorithm was proposed to explore the areas where crime is foreseen according to the crime rate.

2.2. Recent studies on semi-supervised learning and active learning

With the rapid development of data collection from information systems, large-scale data can be obtained from different sources. However, it might be difficult and expensive to acquire a huge labeled data set from the collected data. While supervised learning benefits from the

presence of a large number of labeled data for training, semi-supervised learning utilizes both labeled and unlabeled data to obtain better understanding of population structure. In cases where the amount of labeled data is small, semi-supervised methods are suitable for a better classification. Studies in different domains such as to diagnose skin cancer [13], to classify emotion from speech data [14], and to analyze text sentiments [15], semi-supervised methods are used to handle lack of unlabeled data.

The most popular semi-supervised learning models contain self-learning, multi-view learning, graph-based methods, and mixture models. This study focuses on showing the difference between active learning, self-learning, and random sampling. In the literature, many studies combine self and active learning methods to obtain better results [16].

Although active learning has been applied in various areas such as defect prediction [17], image detection [18], opinion mining [19] and sound classification [20], studies in crime prediction rarely utilized semi-supervised learning. The only study conducted by Nath [21] used expert based semi-supervised learning to select appropriate features.

3. Materials and Methods

Assume that $X = \{x_1, x_2, ..., x_n\}$ is a data set that includes n instances. X is initially divided into training ($X_{Training}$) and test sets (X_{Test}) by the given percentages $k\%$ and $p\%$ of all data, respectively. Random sampling, active learning, and self-learning will be explained in this section using these notations.
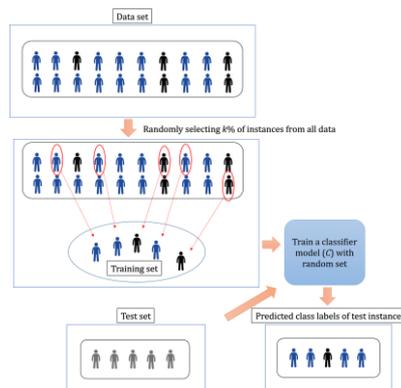
3.1. Random sampling

In the random sampling method, $k\%$ of all the instances in the data set are randomly selected without replacement to be added in $X_{Training}$ and the remaining part is left as X_{Test}. Figure 1 displays the applied procedure. Each instance has an equal probability to be chosen. Using $X_{Training}$, a classifier model is generated to predict the class labels of the samples in X_{Test}. It is the most straightforward and the purest probability sampling strategy.

3.2. Active learning

$X_{Training}$ consists of both labeled and unlabeled instances, described as $X_{Tra\_Labeled}$ and $X_{Tra\_Unlabeled}$. The number of active learning iterations is depicted as I. In each iteration i, the labeled training set $X_{Tra\_Labeled}$ of the relevant iteration is used to construct a classifier model C_i'. Using C_i', the labels of the instances in $X_{Tra\_Unlabeled}$ are predicted. $X_{Tra\_Labeled}$ is then expanded by the selected instances from $X_{Tra\_Unlabeled}$ according to the selection criteria. In this study, entropy-based selection strategy as in Eq. (1) is used, where H_i is the entropy information of an instance in iteration i, P_j is the probability of choice of jth class label, c is the number of classes. After $r\%$ of the unlabeled instances with the highest entropy values are sent to an oracle to be labeled, they are transferred to $X_{Tra\_Labeled}$.

$$H_i = -\sum_{j=1}^{c} P_j log_2 P_j \qquad (1)$$

Figure 1. The general framework of random sampling

This procedure is repeated in each iteration by updating $X_{Tra\_Labeled}$ and $X_{Tra\_Unlabeled}$. The last step is to predict the class label of the instances in X_{Test} using the classifier model C obtained by the expanded training set of $X_{Tra\_Labeled}$. The pseudo-code of the general process is given in Algorithm 1. The advantage of this method is its facility of wisely increasing the labeled training pool in case there is scarce labeled data.

3.3. Self-learning

It is one of the semi-supervised learning algorithms. The difference from active learning is that there is no selection strategy to choose instances from $X_{Tra\_Unlabeled}$ to be included in $X_{Tra\_Labeled}$. Using the initially given $X_{Tra\_Labeled}$, the class labels of all instances in $X_{Tra\_Unlabeled}$ are pseudo-labeled and then all of them with their predicted class labels are added to $X_{Tra\_Labeled}$. As a result, the updated training set is used to reconstruct the classifier model to estimate the labels of the instances in X_{Test}. The general framework is given in Figure 2.

4. Experimental Work

In this study, a number of classification methods (decision tree (DT), naïve Bayes (NB), support vector machines (SVM) and k-nearest neighbours (KNN)) are used as the base learners of three learning techniques (active learning, self-learning, and random sampling). All of the implementations were practiced using Python 3.7.6 on Spyder. The data set used in the experiments, data analysis showing the common statistics in the data and the experimental results of machine learning methods are presented in this section.

4.1. Data set description

Crime data from the police department of England in the year of 2019 are considered in this study. The archive of the past data from 2013 to today is accessible online at the web page of https://data.police.uk/data/archive/.

The features of the data set consist of type ("person search", "vehicle search" or "person and vehicle search"), date of crime case, latitude and longitude information of the crime case, gender of the suspect, age range of the suspect ("under 10", "10-17", "18-24", "25-34", "over 34"), self-defined ethnicity of the suspect, officer-defined ethnicity ("white", "black", "Asian", "mixed", "other"), object of search ("anything to threaten or harm anyone", "article for use in theft", "articles for use in criminal

---

Algorithm 1. Active Learning based on Entropy Information

Inputs:

$X_{Training}$: Initially given training instances which are the combination of $X_{Tra\_Labeled}$ (labeled training instances) and $X_{Tra\_Unlabeled}$ (unlabeled training instances)

$X_{Tra\_Labeled}$final: final training data after active learning iterations

X_{Test}: Test instances whose class labels will be predicted

C_i' : classifier model for the prediction of the class labels of $X_{Tra\_Unlabeled\,i}$ in iteration i using the function f of the selected classifier

C : classifier model for the prediction of the labels of X_{Test} using the function f of the selected classifier

$X_{Tra\_Unlabeled\,ik}$: kth instance of $X_{Tra\_Unlabeled\,i}$ in the ith iteration

m : the number of instances in $X_{Tra\_Unlabeled\,i}$, c : the number of classes

H_i : entropy values of all the instances of $X_{Tra\_Unlabeled\,i}$ in iteration i

Outputs:

P_{ik} : probabilities of class labels for the instances in $X_{Tra\_Unlabeled\,i}$

Y^* : Predicted class labels for the test set

---

for $i = 1$ **to** I **do**
    $C_i' = f(X_{Tra\_Labeled\,i})$
    $H_i = \emptyset$, $H_{i\,\text{Sorted}} = \emptyset$
    for $k = 1$ **to** m **do**
        $P_{ik} = C_i'(X_{Tra\_Unlabeled\,ik})$
        $H_{ik} = -\sum_{j=1}^{c} P_{ik_j}\log_2 P_{ik_j}$
        $H_i = H_i \cup H_{ik}$
    end for
    $H_{i\,\text{Sorted}} = \text{Sort}_{\text{Desc}}(H_i)$
    $X_{Tra\_Unlabeled\,i}' = \{\cup_{z=1}^{t} x_z$ where $x_z \in X_{Tra\_Unlabeled\,i}$ and t denotes the first r% of the instances in $X_{Tra\_Unlabeled\,i}$ according to $H_{i\,\text{Sorted}}\}$
    $X_{Tra\_Labeled\,i+1} = X_{Tra\_Labeled\,i} \cup X_{Tra\_Unlabeled\,i}'$
    $X_{Tra\_Unlabeled\,i+1} = X_{Tra\_Unlabeled\,i} - X_{Tra\_Unlabeled\,i}'$
end for
$C = f(X_{Tra\_Labeled}$final$)$
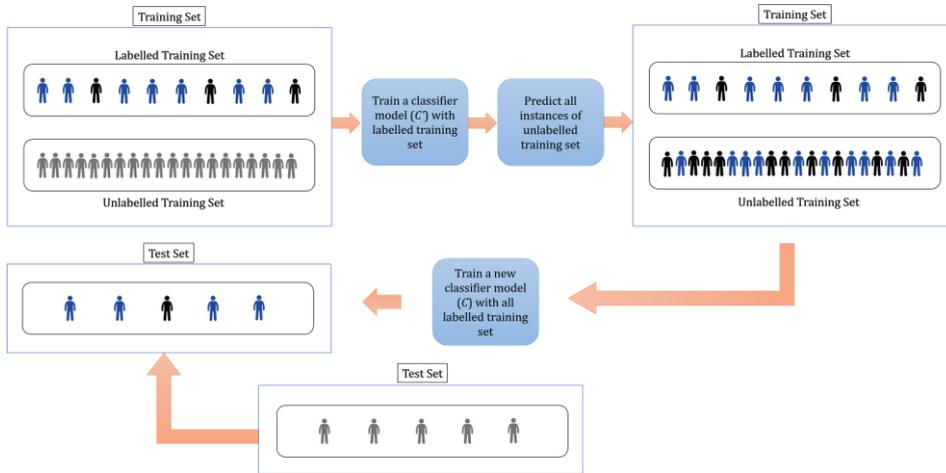$Y^* = C(X_{Test})$

End Algorithm

---

damage", "controlled drugs", "crossbows", "evidence of offences under the act", "evidence of wildlife offences", "firearms", "fireworks", "game of poaching equipment", "goods on which duty has not been paid etc.", "offensive weapons", "psychoactive substances", "stolen

Figure 2. The general framework of self-learning

goods"), outcome, outcome linked to object of search ("Khat or Cannabis warning", "Arrest", "A no further action disposal", "Summons / charged by post", "Penalty Notice for Disorder", "Caution (simple or conditional)", "Community resolution"), city of the event. There are 80362 crime cases recorded for one year.

4.2. Data analysis

In order to present general information about the data, some data analyses were performed. For this purpose, the number of searches according to months, crime searches in terms of gender, and the variation in the number of searches according to crime type were obtained.

Figure 3 demonstrates the map of England in terms of the number of crime searches in 2019. The intensity of the red color indicates the number of cases in a region. According to the figure, Essex was the most searched county while Dorset was the region where crime searches were made at least.

Figure 4 displays the number of searches according to months for the top five crime types. The category of "Controlled drugs" is the most common cause of crime search for all months. "Offensive weapons" and "Article for use in theft" are the other significant causes of searches. Furthermore, while the number of searches decreased in February, the highest number of searches were made in October.

In Figure 5, the effect of gender on the number of searches is shown as a stackplot that the compared to "female". Especially in October, this difference becomes more obvious.

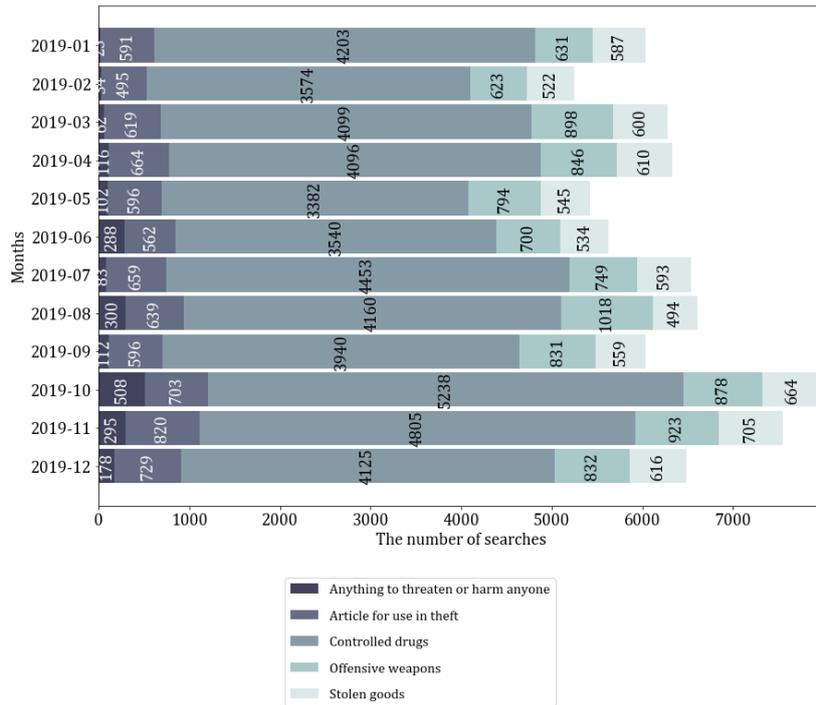**Figure 3.** The number of searches according to the cities of England

Figure 4. The number of searches for crime detection according to months
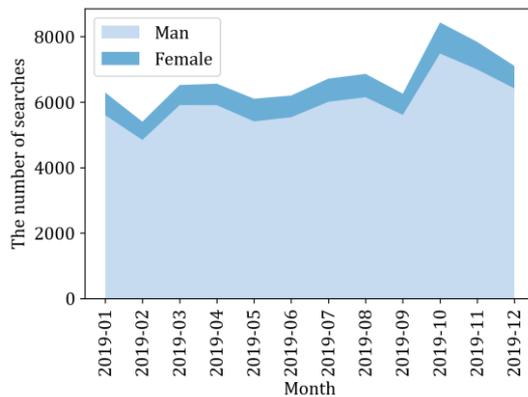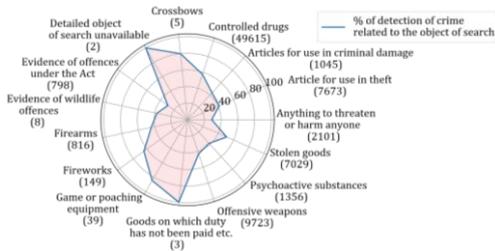

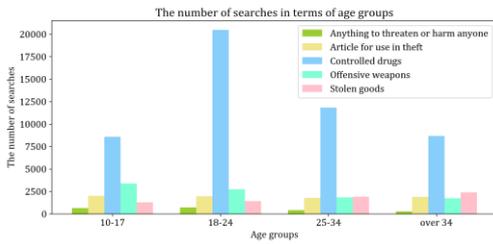**Figure 5.** The number of searches according to gender

The radar chart in Figure 6 is used to compare the detection of crime resulting from the different types of object of searches on a polar grid. The data in this figure was calculated for one year instead of showing for a specific month to observe the general tendency. For example, the total number of crime searches for the category of "Controlled Drugs" is 49615 and 28593 of these suspicious cases were detected as "True" after the decision of the court, so the ratio of the correctly detected crime is 57.6%. The blue line shows this ratio for each category.

The analysis of crime types according to age groups is given in Figure 7. The most common crime type is "Controlled Drugs" for all groups and it reaches the peak point especially in the category of "18-24". As age increases, the rate of "Offensive Weapons" oppositely decreases, and the rate of "Stolen Goods" increases in a parallel way.

Figure 6. The distributions (%) of the criminal cases in terms of the object of the search

Figure 7. The changes in crime type by age

4.3. Comparisons of the methods

Whether the correct outcome of the suspicious event that was searched by the officers was proven as "Crime" or not was predicted using default versions of the base learners, active learning, self-learning, and random sampling in this section. They were compared to each other using four classifiers (decision tree (DT), naïve Bayes (NB), support vector machines (SVM), and k-nearest neighbors (KNN)) as base learner. The performance measures include accuracy, precision, recall, f-score, and the Matthews correlation coefficient (MCC). 10-fold cross-validation was applied in all of the models.

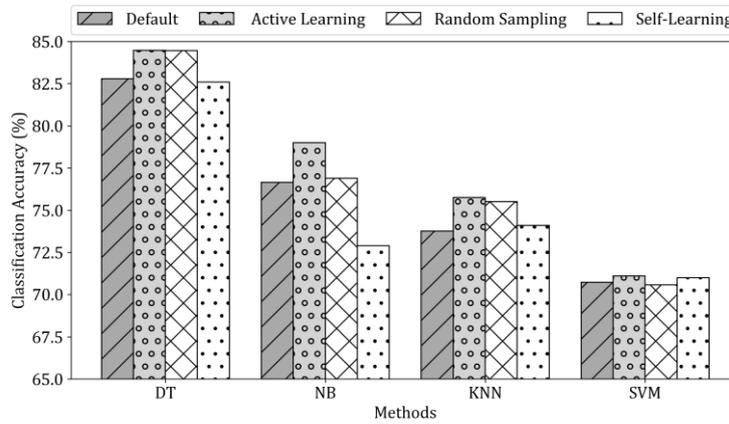There is no parameter tuning applied. The parameters of the base classifiers were selected as default values in the library of the "sklearn" of Python except SVM's kernel function, which was selected as "polynomial" kernel with the degree as 3. SVM was implemented as C-Support Vector Classification based on libsvm. The gamma parameter was set as "scale". Other parameters are as follows. To model DT, "Gini impurity" function was used to measure the quality of a split. "Best split" was used to choose the split at each node. The nodes were expanded until all leaves were pure. Minimum number of samples required to split an internal node was 2. The minimum number of instances needed to be at a leaf node was 1. To generate KNN model, the number of neighbors was 5. The distance metric was Euclidean distance. To compute the nearest neighbors, "auto" method, that decides the most appropriate algorithm based on the values passed to fit method, was selected.

The labeled training data was left as 20% of the whole training data, initially. In active learning, the iterations (I) were determined as 5. In each iteration, 20% of the unlabeled data were labeled using the trained model and then added to the labeled training set. In self-learning, all the unlabeled data were labeled using the model obtained from the labeled data in the beginning and then the labels of the test set were predicted using the model constructed by the combination of both the pre-given labeled data and the data which were labeled later. In random sampling, the selection of instances to the training and test sets were done at random and there is no criterion to choose instances from unlabeled set to the labeled training set.

Initially, total number of samples were 80362. Because 10-fold cross-validation was applied, in each iteration, training and test sets were made up of 72325 and 8037 instances, respectively. While modeling active learning and self-learning, the training set were divided into two parts as labeled and unlabeled sets. 20% of the training set (i.e. 14465 samples) were labeled and the remaining ones (i.e. 57860 samples) were unlabeled. In each active learning iteration, 20% of the unlabeled samples were added to the labeled training set. For example, in the end of the first active learning iteration, 11572 samples (20% of 57860) were moved to the labeled training set so that 26037 samples were obtained as labeled at the beginning of the

increase (%) when active learning is applied instead of default versions of the baseline learners are 0.38, 1.68, 2.35, 1.99 for SVM, DT, NB, and KNN, respectively. This is because active learning smartly selects instances from unlabeled training set (not randomly as in random sampling) and then an expert decides the class label of the selected unlabeled instance. In this way, the number of labeled training samples to generate a classifier are increased in a reasonable way. However, default learners only use the initially given training set which has very limited amount of labeled training samples to model a classifier.

The accuracy values of random sampling are

Figure 8. The comparisons of the applied techniques in terms of accuracy

second active learning iteration. This procedure was repeated in each iteration by expanding the labeled training set.

Figure 8 displays the accuracy results of the applied methods. The results of the active learning method are 84.47, 79.00, 75.75, and 71.11 for DT, NB, KNN, and SVM, respectively. In all cases, active learning outperforms other techniques (especially in NB). The accuracy

84.46, 76.89, 75.51, and 70.58 in the given order in Figure 8. It follows active learning in terms of accuracy, and it is the worst one only in SVM. Self-learning results are 82.60, 72.90, 74.10, and 71.00 for DT, NB, KNN, and SVM, respectively. The reason of the random sampling method to be more successful than self-learning method may be that it uses the real class labels of the instances instead of pseudo-labeled instances.

In general, the accuracy of the applied models when SVM is the base learner is resulted as the lowest one that is because there is no parameter tuning for the applied models, default parameters were used. The identified margin could not distinctly classify the data points. The most accurate predictions are obtained using DT that expresses the given data samples in an optimal way. It is because of its ability of selecting the most discriminatory features. There are some outlier records in the data set. DT requires less data cleaning and is not influenced by outliers, noisy and incomplete data to some fair extent because the leafs of DT are constructed under metrics whose purpose is to discriminate as much as possible the resulting subsets.

Accuracy alone is not enough to measure the performance of a classifier. Other evaluation metrics such as precision (the ratio of correctly predicted positive observations to the total predicted positive observations), recall (the ratio of correctly predicted positive observations to all the observations in actual class), f-score (the harmonic mean of precision and recall) and MCC (a correlation coefficient between the observed and the predicted binary classifications) should also be taken into consideration. Table 1 demonstrates the comparisons of the applied algorithms in terms of the MCC and the weighted versions of precision, recall, and f-score values. Active learning generally achieves the best results of precision, f-score and recall and, only in KNN, random sampling is the optimal one. For binary classifications, MCC generates a high score only if the classifier was able to correctly predict the majority of the positive data instances and the majority of the negative data instances. The best MCC results are obtained with active learning for all learners. So, it is true as in the case of accuracy that a noticeable enhancement is observed when active learning is applied instead of default versions of the baseline learners.

Why random sampling is the closest one to active learning in terms of evaluation metrics can be explained as follows. When applying random sampling, the probability of all instances selected to the training set is equal so that it can provide the classifier with average informative examples. On the other hand, active learning initially generates a model using a training set with limited number of instances. If

Table 1. The comparisons of the applied techniques in terms of precision, recall, f-score, and MCC.

Base Classifier	Method	Precision	Recall	F-Score	MCC
SVM	Active Learning	**71.17**	**71.11**	**71.01**	**0.42**
	Random Sampling	70.58	70.73	70.40	0.41
	Self-Learning	71.08	71.00	70.87	**0.42**
	Default	70.90	70.73	70.53	0.41
DT	Active Learning	**84.48**	**84.47**	**84.47**	**0.69**
	Random Sampling	84.46	84.46	84.46	**0.69**
	Self-Learning	82.60	82.60	82.60	0.65
	Default	82.79	82.79	82.79	0.66
NB	Active Learning	**79.13**	**79.00**	**79.01**	**0.58**
	Random Sampling	77.21	76.89	76.89	0.54
	Self-Learning	73.32	72.94	72.92	0.46
	Default	76.97	76.65	76.65	0.54
KNN	Active Learning	73.91	73.76	73.77	**0.52**
	Random Sampling	**75.64**	**75.51**	**75.52**	0.51
	Self-Learning	74.25	74.08	74.09	0.48
	Default	73.91	73.76	73.77	0.48

this model does not contain enough informative samples, the classifier model will not be able to select good enough samples in the next selection and will achieve a performance close to the average performance achieved by random sampling. The better the initial classifier, the chance of the selection of the most informative instances is high.

5. Conclusion

The aim of this study was to estimate the crime occurrence considering the data of England in 2019. The main approach was handling the situations when there were a few number of labeled samples to construct a classifier model. For this purpose, a semi-supervised learning model (self-learning) and active learning were implemented in addition to simple random sampling. Experiments were done using one of the classifiers (SVM, NB, KNN, DT) as a base learner. According to the results, the most accurate predictions were obtained when active learning was applied.

This study can assist police officers in decision making using similar patterns hidden in past events. Moreover, in case of too much crime searches with unknown outcomes and very little cases with known results, the status of new crime case can be estimated using the applied methodologies (active learning and semi-supervised learning) by expanding the labeled training pool using the pre-given labeled training set.

In the future, spatial and temporal distributions of crime cases can be forecasted using the aforementioned methods and other semi-supervised learning techniques. By changing the percentage of the initially labeled data and the rate of the added unlabeled instances in active learning iterations, the variation in the performance metrics can be observed. Furthermore, ensemble learning strategies may be combined with them in order to make the predictions more accurate.

Acknowledgment

References

[1] Shukla, S., Jain, P.K., Babu, C.R., Pamula, R. 2020. A Multivariate Regression Model for Identifying, Analyzing and Predicting Crimes, Wireless Personal Communications. DOI: 10.1007/s11277-020-07335-w

[2] Kadar, C., Pletikosa, I. Mining Large-Scale Human Mobility Data for Long-Term Crime Prediction 2018. EPJ Data Science, Volume. 7(26). DOI: 10.1140/epjds/s13688-018-0150-z

[3] Agrawal, S., Sejwar, V. 2017. Crime Identification using FP-Growth and Multi Objective Particle Swarm Optimization. IEEE International Conference on Trends in Electronics and Informatics, 11-12 May 2017, Tirunelveli, India, 727-734. DOI: 10.1109/ICOEI.2017.8300799

[4] Nitta, G.R., Rao, B.Y., Sravani, T., Ramakrishiah, N., BalaAnand, M. 2019. LASSO-Based Feature Selection and Naïve Bayes Classifier for Crime Prediction and Its Type, Service Oriented Computing and Applications, Volume. 13(3), p. 187-197. DOI: 10.1007/s11761-018-0251-3

[5] Tayal, D.K., Jain, A., Arora, S., Agarwal, S., Gupta, T., Tyagi, N. 2015. Crime Detection and Criminal Identification in India using Data Mining Techniques, AI & Society, Volume. 30(1), p. 117-127. DOI: 10.1007/s00146-014-0539-6

[6] Shermila, A.M., Bellarmine, A.B., Santiago, N. 2018. Crime Data Analysis and Prediction of Perpetrator Identity using Machine Learning Approach. IEEE 2nd International Conference on Trends in Electronics and Informatics, 11-12 May 2018, Tirunelveli, India, 107-114. DOI: 10.1109/ICOEI.2018.8553904

[7] Khan, J., Lee, Y.K. 2019, LeSSA: A Unified Framework Based on Lexicons and Semi-Supervised Learning Approaches for Textual Sentiment Classification, Applied Sciences, Volume. 9, p. 5562-5590. DOI: 10.3390/app9245562

[8] Iqbal, R., Murad, M.A.A., Mustapha, A., Panahy, P.H.S., Khanahmadliravi, N. 2013. An Experimental Study of Classification Algorithms for Crime Prediction, Indian Journal of Science and Technology, Volume. 6(3), p. 4219-4225. DOI: 10.17485/ijst/2013/v6i3.6

[9] Nguyen, T.T., Hatua, A., Sung, A.H. 2017. Building a Learning Machine Classifier with Inadequate Data for Crime Prediction. Journal of Advances in Information Technology, Volume. 8(2), p. 141-147. DOI: 10.12720/jait.8.2.141-147

[10] Alves, L.G., Ribeiro, H.V., Rodrigues, F.A. 2018. Crime Prediction through Urban Metrics and Statistical Learning. Physica A: Statistical Mechanics and its Applications, Volume. 505, p. 435-443. DOI: 10.1016/j.physa.2018.03.084

[11] Chhabra G., Vashisht V., Ranjan J., Crime Prediction Patterns Using Hybrid K-Means Hierarchical Clustering. Journal of Advanced Research in Dynamical Control Systems, Volume. 11, p. 1249-1258

[12] Sivanagaleela B., Rajesh S. 2019. Crime Analysis and Prediction using Fuzzy C-Means Algorithm. IEEE 3rd International Conferences on Trends in Electronics and Informatics (ICOEI), 23-25 April, Tirunelveli, 595-599. DOI: 10.1109/ICOEI.2019.8862691

[13] Masood, A., Al-Jumaily, A., Anam, K. 2015. Self-Supervised Learning Model for Skin Cancer Diagnosis. 7th International IEEE/EMBS Conference on Neural Engineering (NER), 22-24 April, Montpellier, 1012-1015. DOI: 10.1109/NER.2015.7146798

[14] Esparza J., Scherer S., Schwenker F. 2012. Studying Self- and Active-Training Methods for Multi-Feature Set Emotion Recognition. pp 19-31. Schwenker F., Trentin E., ed. 2011. Partially Supervised Learning. PSL 2011. Lecture Notes in Computer Science, vol 7081. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-28258-4_3

[15] Khan, J., Lee, Y.K. 2019. LeSSA: A Unified Framework Based on Lexicons and Semi-Supervised Learning Approaches for Textual Sentiment Classification. Applied Sciences, Volume. 9(24). DOI: 10.3390/app9245562

[16] Karlos, S., Kanas, V.G., Aridas, C., Fazakis, N., Kotsiantis, S. 2019. Combining Active Learning with Self-train algorithm for Classification of Multimodal Problems. IEEE 10th International Conference on Information, Intelligence, System and Applications (IISA), 15-17 July, Patras, Greece, 1-8. DOI: 10.1109/IISA.2019.8900724

[17] Li, F., Qu, Y., Ji, J., Zhang, D., Li, L. 2020. Active Learning Empirical Research on Cross-Version Software Defect Prediction Datasets. International Journal of Performability Engineering, Volume. 16(4), p. 609-617. DOI: 10.23940/ijpe.20.04.p12.609617

[18] Kellenberger, B., Marcos, D., Lobry, S., Tuia, D. 2019. Half a Percent of Labels is Enough: Efficient Animal Detection in UAV Imagery Using Deep CNNs and Active Learning. IEEE Transactions on Geoscience and Remote Sensing, Volume. 57(12), p. 9524-9533. DOI: 10.1109/TGRS.2019.2927393

[19] Vitório, D., Souza, E., Oliveira, A.L. 2019. Using Active Learning Sampling Strategies for Ensemble Generation on Opinion Mining. 8th Brazilian Conference on Intelligent Systems (BRACIS), 15-18 October, Salvador, Brazil, 114-119. DOI: 10.1109/BRACIS.2019.00029

[20] Han, W., Coutinho, E., Ruan, H., Li, H., Schuller, B., Yu, X., Zhu, X. 2016. Semi-Supervised Active Learning for Sound Classification in Hybrid Learning Environments. PloS one, Volume. 11(9). DOI: 10.1371/journal.pone.01620

[21] Nath, S.V. 2006. Crime Pattern Detection Using Data Mining. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops, 18-22 December, Hong Kong, China, 41-44. DOI: 10.1109/WI-IATW.2006.55