Environment Protection Engineering

Vol. 47 2021 No. 4

DOI:10.37190/epe210404

ROBERT GIEL (ORCID0000-0002-7406-0044)[1]
ALICJA DĄBROWSKA (ORCID0000-0003-2526-6916)
SYLWIA WERBIŃSKA-WOJCIECHOWSKA (ORCID0000-0002-9228-982X)

ACTIVE LEARNING FOR AUTOMATIC CLASSIFICATION OF COMPLAINTS ABOUT MUNICIPAL WASTE MANAGEMENT

Information flow is an important issue in the area of waste management. There is a need for a fast response to reported problems. Therefore we investigated the classification process of Polish waste-related complaints sent by Wrocław's residents. It has been noticed that residents, mostly without expert knowledge of waste management, incorrectly classify the observed problems. In response to the observed unacceptable classification accuracy, we introduced a multi-class machine learning classification. Machine learning is widely used in waste management issues like predicting waste generation or different waste fractions identification for automated sorting. However, based on the literature review, it can be stated that there is a lack of solutions in machine learning-based text classification regarding waste management. Ten chosen classifiers were used to classify considered complaints into defined categories automatically. Additionally, we incorporated the active learning approach to reduce experts' effort involved in the labeling process, which is necessary when having an unlabeled dataset. The results confirm the possibility of applying machine learning algorithms to waste-related Polish complaints.

1. INTRODUCTION

Due to the rapid development of cities and the growing population of societies, one of the global problems is ensuring effective waste management [1]. This requires introducing proper ways of, e.g., planning, collecting, segregation, or waste recycling/disposal [2]. However, following the accompanying document *Early Warning Report for Poland on the Implementation of EU Waste Legislation* [3], Poland is one of 18 EU

[1]Faculty of Mechanical Engineering, Department of Operation and Maintenance of Technical Systems, Wrocław University of Science and Technology, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland, corresponding author A. Dąbrowska, email address: alicja.dabrowska@pwr.edu.pl

countries being at risk of non-compliance with the 2025 target for recycling of municipal waste set out in Article 11(2)a of Directive 2008/98/EC [4]. In one of the biggest Polish cities, Wrocław, located in the Lower Silesia region, achieving the required level of recycling was fulfilled through the years 2012–2018. However, considering the indicator growth rate, there is a threat of not reaching the required level in future years. This means that municipal companies and local governments' recent decisions in the field of waste management are insufficient.

The conducted research analysis of municipal waste management systems in a chosen Polish city, Wrocław, showed that one of the main problems is connected with an information management system. On the one side, there is a necessity to provide a non-disturbed information flow in a complex and diverse business environment with different types of stakeholders involved. On the other, the requirement of reliable and fast response for residents' complaints about waste services performance cannot be fully satisfied. This is mainly connected with long delays in residents' complaints responses due to the incorrect categorization of complaints. Following this, it is justifiable to introduce an automatic classification of residents' complaints according to predefined categories adjusted to the municipal waste management systems requirements.

Automatic text classification is a research area where many interesting problems and applications have emerged in recent years. This interest is primarily driven by the rapid development of the Internet and access to many different sources of information (online databases), resulting in the need for appropriate classification and analysis of the acquired data [5]. When having access to large-scale information, manual classification by domain experts becomes ineffective and unfeasible. As a result, we may find many research works focused on classification methods based on supervised (e.g., [6]) or unsupervised learning (e.g., [7]). Currently, there are also many different applications of text classification methods, e.g., in production research (e.g., [8]) or social networks/media (e.g., [9]). Moreover, many research works focused on the automatic classification of text written in different languages/dialects (see, e.g., [10]).

Moreover, a review of recent literature reveals an increasing number of studies covering many different issues and problems of waste management concerning the implementation of machine learning (ML) tools. The obtained solutions are compatible with the main waste management process stages and are dedicated to the two main areas – material flows management and information flow management.

The material flows management is strictly connected with the issues of providing a performance of physical transport, collection, transshipment, and processing of waste. The proposed machine learning solutions in this area mostly regard the problems of adequate collection and segregation technologies development (see, e.g., [11–13]). Information flow management is mostly dedicated to the solutions that give the possibility for decision-making process performance. The proposed machine learning solutions in this area are mostly focused on providing reliable data for planning waste collection (e.g., [14]) or identification and tracking of bins/waste or vehicles (e.g., [15]). Many solutions are also dedicated to sustainable development providing (e.g., [16–18]).

Despite the growing interest in the use of ML algorithms in the area of text classification, there is a lack of solutions dedicated to waste information management in the context of the implementation of text classification that would take into account its specific requirements connected with, e.g., necessary reaction speed, the occurrence of different types of reporting users with different levels of waste-related knowledge, or waste language specificity. Therefore, this research's primary intention is to develop an automatic text classification of complaints about municipal waste management written in Polish. The proposed automatic classification of complaint reports is based on four main steps: data pre-processing, feature extraction, active learning-based classification, results. The classification was performed according to the company's categories, such as waste collection, winter-connected problems, waste containers/bags, waste segregation, city cleaning, declarations, administrator zone, and others.

2. MATERIALS AND METHODS

Data characteristics. The dataset used for the classification contains complaints sent by residents of Wrocław. The analyzed complaints regard observed waste-related problems.

Currently, in Poland, responsibility for the proper management of collecting and processing waste corresponds to the municipality. These municipalities are obligated to select special purpose companies responsible for the process of municipal waste company selection and supervision. The chosen companies are required to achieve the stated objectives, including ensuring the city's cleanliness, snow cleaning, waste collection, and processing.

Since August 2017, the company supervising waste management in the city of Wrocław has introduced a complaint reporting regarding waste-related issues (Fig.1). Residents can choose one of eight categories, as follows: waste collection, winter-connected problems, waste containers/bags, waste segregation, city cleaning, declarations, administrator zone, and others. Some of them have subcategories.

Employees of the company's departments later analyze the complaints recorded in the system and contact the appropriate subcontractor for intervention. Incorrect category assignment to the complaint by a resident may lead to a longer response time for the company. This is mainly connected with the necessity of sending messages between individual departments in the company.

From August 2017 to April 2019, the company received about 24 539 complaints (about 1169 reports per month). Over the years, some categories have been removed or modified. There can be defined three types of reporting users:

• employees of the company (42% of complaints) – expert knowledge in waste management, responsible for the classification of all the complaints reported by phone or email,

• residents – (32% of complaints) – in most cases without expert knowledge in waste management,

• administrators – (21% of complaints) – a separate group of residents responsible for reporting only one type of waste-related problem (category Administrator zone with two subcategories: high volume waste and green waste).

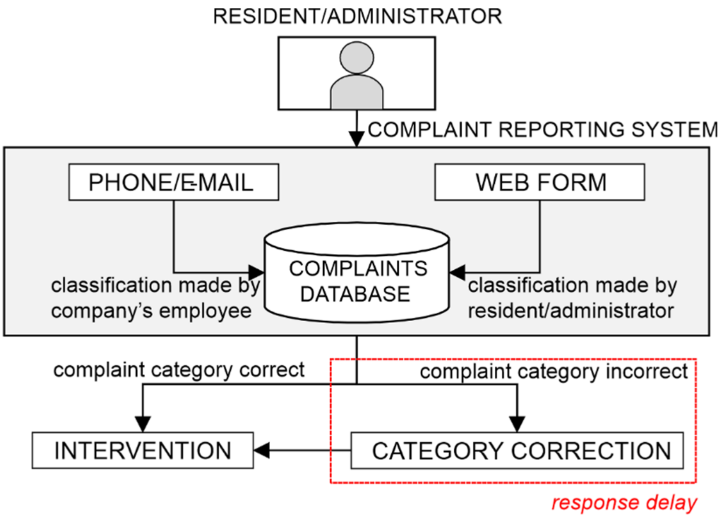

Fig. 1. Complaint reporting system

Each complaint is stored in a database. The main information related to each record includes the address to which the complaint relates, apartment number, sector (the city is divided into four sectors for waste collection purposes), housing estate, category, subject, description, name of the applicant, date of creation of the application and date of closing the application. The description contains the information necessary to identify the problem and perform appropriate actions. Therefore, on its basis, it is possible to identify to which category it should be classified (which one is the most suitable based on the recorded data). The description contains an average of 38 words, with a standard deviation of 31. The maximum number of detected words per application is 667.

The dataset used for classification purposes consists of complaints and categories assigned to them. However, it has been noticed that some complaints are classified incorrectly. Based on a randomly selected sample (2400 complaints), the analysis of correctly classified notifications was made, taking into account each reporting user type. Obtained classification Accuracy (number of correctly classified complaints divided by all classified complaints) considering types of reporting users is:

- company's employees – 94.30%,
- residents – 84.37%,
- administrators – 98.05%.

Further analysis showed that the most common mistakes were made under the category marked as Other. This gives the authors the possibility to assume that it was the category chosen when the user had a problem with choosing the appropriate category of his or her complaint.

Based on the obtained classification accuracy, it was assumed that the correctness obtained in the case of two types of reporting users (administrators and the company's employees) is at a satisfactory level. The decision was based on the assessment that the acceptance level of classification accuracy is 90%. Based on the limit value of 90%, only one group of complaint reports obtained from residents is unacceptable. Following this, the authors focused their attention only on this type of reporting users. Additionally, it should be noticed that due to the observed incorrect classification, the considered dataset should be treated as an unlabeled dataset. For this reason, a group of experts is required for the data pre-processing stage, where complaints must be labeled by assigning appropriate categories.

Active learning-based automatic text classification for waste-related complaints. To improve residents' classification accuracy, it was proposed to modify the reporting system by introducing automatic classification of complaint reports sent by residents.

Within the framework of carrying out automatic text classification and verification of the ability to improve the current system performance, it is necessary to perform four main steps: data pre-processing, feature extraction, classification with the use of chosen classifiers, and evaluation. The machine learning-based classification was carried out following the active learning approach.

The first step regards data pre-processing, in which the goal is to remove some of the test entries and entries without description. In addition, Polish characters were also removed. The initial set of notifications containing 24 539 was reduced to 23 466.

The analyzed dataset contains complaints consisting of a description and the category assigned to it. Due to detected errors in the assigned categories, the dataset has to be treated as an unlabeled dataset, and it is necessary to carry out manual labeling by a group of experts. 2000 complaints (basic set) were drawn according to the uniform distribution for their use in the learning and testing process. An additional 400 complaints (validation set) covering the first three weeks of March 2019 were selected for validation process performance. The introduction of automatic classification is to be dedicated to one group of reporting groups – residents; hence the prepared sample for validation contains descriptions of reports only from this group of users.

A randomly chosen basic set containing 2000 complaints is a good representation of the entire dataset. The percentage of reports in the defined categories is similar in both sets. In general, complaints within its categories are repetitive, including their form and content. Due to this, we decided to use only chosen part of the entire dataset.

In the case of the system analyzed in this article, it was necessary to separate an additional set of validation data to check classifiers' performance for a real-life system. Considering the data from the last month being available in the dataset gives the possibility to reflect the operation of the developed method of automatic classification in real conditions. Figure 2 presents the selected sample division into sets that are used for training, testing, and validation.

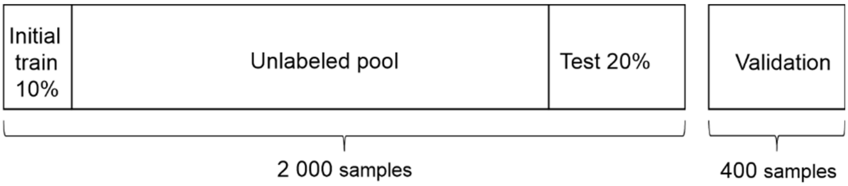

Fig. 2. The classification of the selected sample of data

In the developed model of automatic classification of complaint reports, the authors used one of the feature extraction techniques for text data – Bag of Words with term-frequency-inverse document frequency (TF-IDF). ML algorithms require the transformation of used text data into a vector form [19]. Thanks to the Bag of Words strategy implementation, complaints are perceived through an individual word occurrence prism. According to this approach, tokenization and word occurrence counting in documents /applications were performed as a part of feature extraction. In addition, TF-IDF was introduced to determine the weight of individual words. The TF-IDF approach implementation allows reflection of the relevance of individual words. Thus, words' frequency is taken into account and their relationship to the topic [20]. Thanks to this, each application's most frequently occurring words receive less weight than the unique words that determine membership in a given class.

Moreover, the authors used ten classifiers to automatically classify selected complaints: Random forest (RF), Nearest neighbors (KNN), Multinomial naïve bayes (MNB), Decision tree (DT), Bernoulli naïve bayes (BNB), AdaBoost (AB), Logistic regression (LR), Support vector classifier with linear kernel (SVC1), Support vector classifier with Gaussian kernel (SCV2), and Support vector classifier with Sigmoid kernel (SVC3). Their detailed descriptions can be found, e.g., in [21, 22].

Regardless of the classifier used, it is necessary to carry out the learning and testing process. Thus, the use of ML algorithms for automatic classification is connected with the necessity of labeled dataset use. Without properly assigned labels to the analyzed data, it is impossible to carry out the learning process and the testing process, i.e., the two critical processes of supervised learning. Most publications are available in this field present solutions based on labeled datasets. Therefore, the most commonly used process is passive learning (PL), in which the learner receives a ready-to-learn data set and does not interfere in its preparation. This type of learning is often associated with using an expert/group of experts to carry out manual labeling. However, in the case of large datasets, this process may be too time and/or cost-consuming.

The indicated problematic nature of the PL approach, when it is necessary to classify unlabeled dataset, has led to an alternative approach, i.e., active learning (AL). AL is a subset of ML which allows active participation of the algorithms in the labeling process. Lerner is taking part in data annotation, leading to better classification performance with fewer training samples. In the case of limited time, human resources, and/or a budget, reducing the sample required for manual labeling can significantly facilitate the automatic classification process.

Following this to carry out the learning process following the active approach, the 2000 complaints were divided into three parts for learning and testing (according to Fig. 2):

- initial training set – 10% of the complaints, randomly selected according to the uniform distribution,
- unlabeled pool – 70% of the reports, set of complaints without labels,
- test set – 20% of the reports.

The use of AL requires making decisions primarily on how to create the initial train set, choose a query strategy (strategy for selecting samples to be labeled), and define stopping criteria. Based on the conducted literature review, it was decided to:

- use a random selection of the initial train set,
- use uncertainty sampling (based on a margin of confidence) as query strategy, and
- determine the number of iterations after which the AL process is stopped.

The general scheme of the applied active learning approach is shown in Fig. 3.

A learning process is carried out using the selected classifiers' initial training set after conducting text pre-processing and feature extraction. Then the given classifier is tested, followed by an assessment of its accuracy and other metrics. The condition which leads to stopping the learning process is achieving 140 iterations. During each iteration, the unlabeled pool samples are ranked based on the difference between the top two most confident predictions of the category calculated within each sample. There is a package of 10 samples with the smallest margin of confidence achieved from the unlabeled pool set. This package is later labeled by an expert, after which it is removed from the unlabeled pool and used to increase the Initial training set. With the new learning sample prepared in this way, the learning, testing, and assessment process are carried out again. As a result of AL, it is possible to check at which size of a training set the assessment ratio level is satisfactory.

To assess and select the best classifiers for the considered dataset, we have chosen a few metrics commonly used in text classification for multi-class categorization [23]. All those metrics are defined employing the following features:

- true positive *TP* – observations correctly predicted as a positive class,
- true negative *TN* – observations correctly predicted as a negative class,

• false positive *FP* – observations incorrectly assigned to the positive class, which belong to the negative class,

• false negative *FN* – observations incorrectly assigned to the negative class, which belong to the positive class.

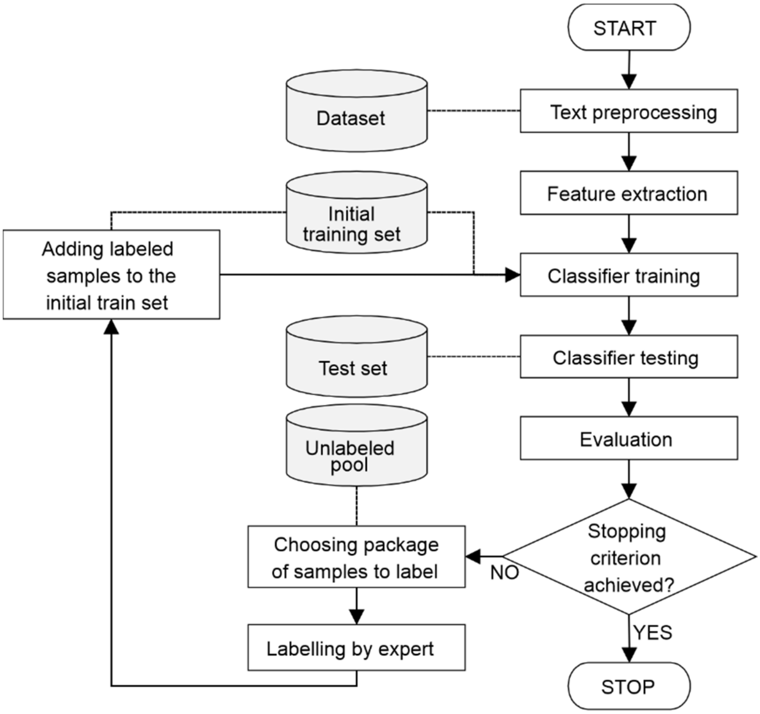

Fig. 3. General scheme of the applied active learning approach

Presented features are calculated for every class in the case of multi-class classification. Considering k classes, the number of observations in every class n_i and number of TP, TN, FP, and FN for every class, Accuracy (A), weighted precision (P_w), weighted recall (R_w) and weighted F1 score (F_{1w}) can be defined as [24]:

$$A = \frac{\sum_{i=1}^{k} \frac{TP_i + TN_i}{TP_i + FP_i + FN_i + TN_i}}{k} \tag{1}$$

$$P_w = \frac{\sum_{i=1}^{k} \frac{TP_i}{TP_i + FP_i}}{\sum_{i=1}^{k} n_i} \tag{2}$$

$$R_w = \frac{\sum_{i=1}^{k} \frac{n_i TP_i}{TP_i + FN_i}}{\sum_{i=1}^{k} n_i} \tag{3}$$

$$F_{1w} = \frac{\sum_{i=1}^{k} \frac{2 n_i TP_i}{2TP_i + FP_i + FN_i}}{\sum_{i=1}^{k} n_i} \tag{4}$$

There are three ways for calculating P_w, R_w, and F_{1w}: macro-averaged (metrics are calculated for every class, but do not consider class imbalance), micro-averaged (metrics are calculated globally) and weighted averaged (metrics are calculated for every class with class imbalance consideration). We decided to use weighted averaged metrics to evaluate the performance of the chosen classifiers due to the existence of class imbalance.

3. RESULTS

Table 1 presents a comparison of the accuracy values obtained in the next iterations (for different training set sizes) using the train/test split approach and basic set and the use of train/test split approach and validation set as a test set. It can be noticed that with the use of a basic set, as many as 7 out of 10 algorithms had a better match than for the classification of complaints made by residents. The SVC1 algorithm achieved the highest result: 91.00%, which is an improvement of 6.63% (2.79% for the whole system).

Table 1

Active learning classification; accuracy for basic and validation sets [5]

Algorithm	Accuracy for basic set (1)	Train set size (% of 2000)	Accuracy for validation set (2)	(1) – (2)
RF	89.75	46.00	87.78	1.97
KNN	84.25	23.50	81.80	2.45
MNB	89.00	32.50	84.54	4.46
DT	81.75	8000	77.81	3.94
BNB	90.25	21.50	85.29	4.96
AB	73.00	16.00	73.07	-0.07
LR	89.75	42.00	85.79	3.96
SVC1	91.00	69.00	86.78	4.22
SVC2	89.00	63.50	86.78	2.22
SVC3	90.00	31.00	87.28	2.72

For the validation set, 7 of the tested algorithms obtained a better result than the residents. In this case, RF algorithm was the best and allowed for an improvement of 3.41% (1.39% for the entire system).

Tables 2 and 3 present the results of the precision, recall, and $F1$score metrics assessment. The results are given for basic and validation test sets. The precision metric is slightly lower than obtained accuracy for the basic set.

Table 2

Active learning classification; precision, recall and F1 score for basic set [%]

Algorithm	Train set size (% of 2000)	Precision	Recall	F1 score
RF	46.00	87.98	89.75	88.22
KNN	23.50	84.24	84.25	83.94
MNB	32.50	85.19	89.00	86.90
DT	80.00	79.96	81.75	80.49
BNB	21.50	89.98	90.25	88.15
AB	16.00	74.72	7300	71.38
LR	42.00	88.49	89.75	87.46
SVC1	69.00	89.08	91.00	89.85
SVC2	63.50	84.28	89.00	86.52
SVC3	31.00	88.54	90.00	89.08

Table 3

Active learning classification; precision, recall and F1 score for validation set [%]

Algorithm	Precision	Recall	F1 score
RF	87.98	89.75	88.22
KNN	84.24	84.25	83.94
MNB	85.19	89.00	86.90
DT	79.96	81.75	80.49
BNB	89.98	90.25	88.15
AB	74.72	73.00	71.38
LR	88.49	89.75	87.46
SVC1	89.08	91.00	89.85
SVC2	84.28	89.00	86.52
SVC3	88.54	90.00	89.08

The learning speed of the three best classifiers, SVC1, BNB, and SVC3, for the basic and validation sets, are presented in Figs. 4–6.

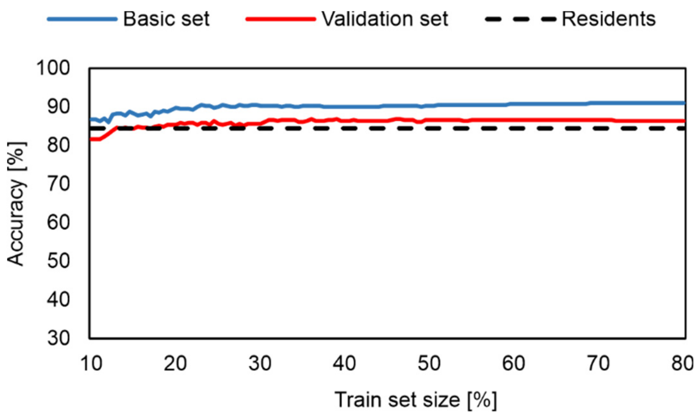

Fig. 4. The learning speed for SVC1 classifier

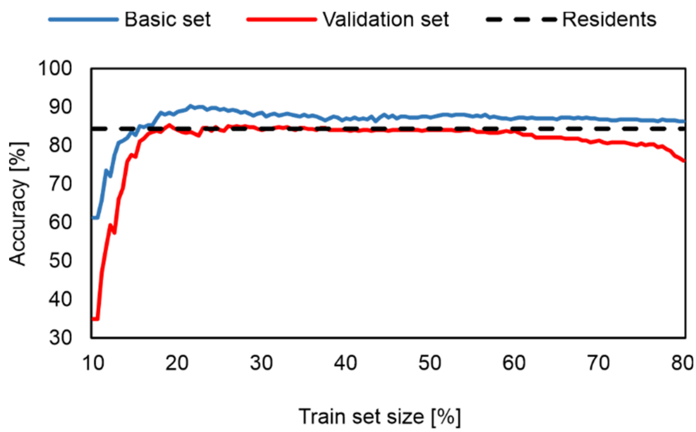

Fig. 5. The learning speed for BNB classifier

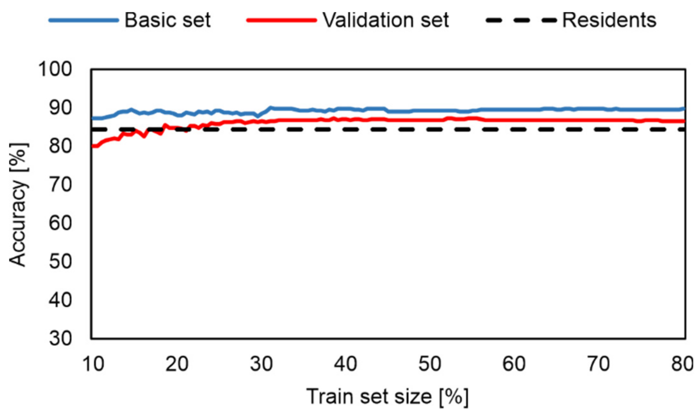

Fig. 6. The learning speed for SVC3 classifier

The accuracy obtained based on SVC algorithms using only 10% training size was characterized by an adequate matching level. With 15–20% train set size, they achieved greater accuracy than residents' classification. In the case of the BNB algorithm implementation, despite achieving one of the highest accuracy levels for 19–21% of the sample size, the values continued to oscillate at the limit of residents' accuracy (for the validation sample). Thus, this algorithm is less preferred, especially since after using 50% train size, there is a noticeable decrease in obtained accuracy level.

It is difficult to answer why there were obtained lower Accuracy levels for the validation set at this stage. However, it is to be expected that this may be affected by so-called data seasonality. Moreover, in one of the previously published articles [25], we showed a significant increase in waste selection complaints when the contractor for this service had changed. Therefore, it is also necessary to check the influence of external factors (e.g., time of the year, contractors change) on the level of evaluated metrics for the selected algorithms.

4. CONCLUSIONS AND DIRECTIONS FOR FURTHER RESEARCH

In the presented paper, we introduced a method for analyzing and classifying residents' complaints reported in a chosen municipal waste management system. The analyzed complaint letters are written in Polish and can be classified into such categories as waste collection, winter-connected problems, waste containers/bags, waste segregation, city cleaning, declarations, administrator zone, and others. The proposed approach was based on active learning implementation. The classification experiments were performed using ten ML classifiers.

Additionally, it was indicated that it is necessary to introduce a validation set (latest data from datasets) when the analyzed dataset is related to the information system with the seasonality of data. This allows checking the effectiveness of the proposed model in real conditions because achieving satisfactory evaluation results in the testing process is not equivalent to achieving the same results after introducing automatic classification into the real system.

Based on the obtained results, it is possible to conclude that the automatic classification based on machine learning (ML) algorithms proposed in this paper can successfully replace the traditional classification method used for one type of reporting user – residents. In the case of the other types of reporting users, employees of the municipal company (who should be treated as experts) and administrators (who only report the need for one of the two types of containers), this method is not sufficient enough.

The application of the ML-based method in the area of automatic classification of residents' requests will allow, on the one hand, simplifying the process of complaint reporting, and on the other hand, reducing the number of misclassified complaints, which may translate into shorter response time to the reported problem. To determine the best classifier for the analyzed case, four basic measures were analyzed – accuracy, precision, recall, and F1 score. The obtained results indicated that the best classifiers for the analyzed classification problem are SVC algorithms.

In conclusion, the obtained results confirm the usefulness of the application of automatic classification for text data sets in the form of complaints related to the subject of waste management (causing difficulties in classification due to a small number of unique words within classes) and characterized by the presence of the Polish language (problematic in classification operations).

At the same time, it is essential to note that the validation process was conducted using complaints reported during the three weeks of March. However, the waste collection and management system is susceptible to seasonal changes and external factors such as subcontractor change. Therefore, further research analysis should be focused on investigating their impact on the classification process's accuracy. This will allow for a better understanding of the dependencies within the system and the development of possible changes to the presented approach to achieve an even higher level of accuracy.

REFERENCES

[1] Pasiecznik I., Banaszkiewicz K., Syska Ł., *Local community e-waste awareness and bahavior. Polish case study*, Environ. Prot. Eng., 2017, 43 (3), 287–303. DOI: 10.37190/epe170320.

[2] Jakubus M., Stejskal B., *Municipal solid waste management systems in Poland and the Czech Republic. A comparative study*, Environ. Prot. Eng., 2020, 46 (3), 61–78. DOI: 10.37190/epe200304.

[3] Sarc R., Curtis A., Kandlbauer L., Khodier K., Lorber K.E., Pomberger R., *Digitalisation and intelligent robotics in value chain of circular economy oriented waste management. A review*, Waste Manage., 2019, 95, 476–492. DOI: 10.1016/j.wasman.2019.06.035.

[4] *The Early Warning report for Poland Accompanying the document.* Report from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the implementation of EU waste legislation, SWD (2018) 426 final. Brussels. DOI: 10.1017/CBO9781107415324.004.

[5] Directive (EU) 2018/850 of the European Parliament and of the Council of 30 May 2018 amending Directive 1999/31/EC on the landfill of waste. Retrieved from http://data.europa.eu/eli/dir/2018/850/oj.

[6] Montebruno P., Bennett R.J., Smith H., van Lieshout C., *Machine learning classification of entrepreneurs in British historical census data*, Inf. Proc. Manage., 2020, 57 (3), 1–22. DOI:10.1016/j.ipm.2020.102210.

[7] HaCohen-Kerner Y., Dilmon R., Hone M., Ben-Basan M.A., *Automatic classification of complaint letters according to service provider categories*, Inf. Proc. Manage., 2019, 56 (6), 1–20. DOI: 10.1016/j.ipm.2019.102102.

[8] Ko Y., Seo J., *Text classification from unlabeled documents with bootstrapping and feature projection techniques*, Inf. Proc. Manage., 2009, 45 (1), 70–83. DOI: 10.1016/j.ipm.2008.07.004.

[9] Zhou Y., Yang S., Li Y., Chen Y., Yao J., Qazi A., *Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining*, Inf. Proc. Manage., 2020, 57 (2), 1–11. DOI: 10.1016/j.ipm.2019.102179.

[10] Kozlowski D., Lannelongue E., Saudemont F., Benamara F., Mari A., Moriceau V., Boumadane A., *A three-level classification of French tweets in ecological crises*, Inf. Proc. Manage., 2020, 57 (5), 1–20. DOI: 10.1016/j.ipm.2020.102065284.

[11] LI L., LI W., GONG D., *Naive Bayesian automatic classification of railway service complaint text based on eigenvalue extraction*, TV-TG, 2019, 26 (3), 778–785. DOI: 10.17559/TV-20190420161815.

[12] BAGHERI M., ESFILAR R., GOLCHI M.S., KENNEDY C.A., *A comparative data mining approach for the prediction of energy recovery potential from various municipal solid waste*, Renew. Sustain. Energy Rev., 2019, 116, 1–12. DOI: 10.1016/j.rser.2019.109423.

[13] FERRER J., ALBA E., BIN-CT: *Urban waste collection based on predicting the container fill level*, Biosyst., 2019, 186, 1–9. DOI: 10.1016/j.biosystems.2019.04.006.

[14] PRASANNA M.A., VIKASH KAUSHAL S., MAHALAKSHMI P., *Survey on identification and classification of waste for efficient disposal and recycling*, Int. J. Eng. Technol., 2018, 7 (2.8), 520–523. DOI: 10.14419 /ijet.v7i2.8.10513.

[15] KANNANGARA M., DUA R., AHMADI L., BENSEBAA F., *Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches*, Waste Manage., 2018, 74, 3–15. DOI: 10.1016/j.wasman.2017.11.057.

[16] DIMILILER K., EVER Y.K., MUSTAFA S.M., *Vehicle detection and tracking using machine learning techniques*, 10th Int. Conf. Theory and Application of Soft Computing with Words and Perceptions – ICSCCW, 2020, 373–381. 10.1007/978-3-030-35249-3_48.

[17] MALINAUSKAITE J., JOUHARA H., CZAJCZYŃSKA D., STANCHEV P., KATSOU E., ROSTKOWSKI P., THORNE R.J., COLON J., PONSA S., AL-MANSOUR F., ANGUILANO L, KRZYŻYŃSKA R., LOPEZ J.C., VLASOPOULOS A., SPENCER N., *Municipal solid waste management and waste-to-energy in the context of a circular economy and energy recycling in Europe*, Energy, 2017, 141, 2013–2044. DOI: 10.1016/j.energy.2017.11.128.

[18] ZHANG A., VENKATESH V.G., LIU Y., WAN M., QU T., HUISINGH D., *Barriers to smart waste management for a circular economy in China*, J. Clean. Prod, 2019, 240, 1–12. DOI: 10.1016/j.jclepro. 2019.118198.

[19] ALLAHYARI M., POURIYEH S., ASSEFI M., SAFAEI S., TRIPPE E.D., GUTIERREZ J.B., KOCHUT K., *A brief survey of text mining: Classification, clustering and extraction techniques*, ArXiv, 2017, 1–13. DOI: http://arxiv.org/abs/1707.02919.

[20] DZISEVIC R., SESOK D., *Text classification using different feature extraction approaches*, Open Conference of Electrical, Electronic and Information Sciences, eStream 2019, Proc. IEEE, 1–4. DOI: 10.1109/eStream.2019.8732167.

[21] KADHIM A.I., *Survey on supervised machine learning techniques for automatic text classification*, Artif. Intell. Rev., 2019, 52 (1), 273–292. DOI: 10.1007/s10462-018-09677-1.

[22] KOWSARI K., MEIMANDI K.J., HEIDARYSAFA M., MENDU S., BARNES L.E., BROWN D.E., *Text classification algorithms: A survey*, Information (Switzerland), 2019, 10 (4), 1–68. DOI: 10.3390/info10040150.

[23] SOKOLOVA M., LAPALME G., *A systematic analysis of performance measures for classification tasks*, Inf. Proc.. Manage., 2009, 45 (4), 427–437. DOI: 10.1016/j.ipm.2009.03.002.

[24] BEHERA B., KUMARAVELAN G., KUMAR P., *Performance evaluation of deep learning algorithms in biomedical document classification*, 11th Int. Conf. Advanced Computing (ICoAC), Chennai, India, 2019, 2020–2024. DOI: 10.1109/ICoAC48765.2019.246843.

[25] DĄBROWSKA A., GIEL R., PLEWA M., *Assessment of the municipal waste collection process on the basis of irregularities notifications sent by residents. A case study in Wroclaw, Poland*, Environ. Prot. Eng., 2020, 46 (2), 79–92. DOI: 10.37190/epe200206.