

Finding Anomalies in Waste Transportation Data with Supervised Category Models

António Pereira Barata^{1,2,3}[0000–0002–0540–7681], Gerrit Jan de Bruin^{1,2,3},
Frank Takes³, Jaap van den Herik¹, and Cor Veenman^{3,4}

¹ Leiden Centre of Data Science, Leiden University, Netherlands

² Human Environment and Transport Inspectorate, Ministry of Infrastructure and Water Management, Netherlands

³ Leiden Institute of Advanced Computer Science, Leiden University, Netherlands

⁴ TNO, Data Science Department, the Hague, Netherlands
`a.p.pereira.barata@liacs.leidenuniv.nl`

Abstract. Because of environmental risks, the transportation of waste materials within the EU is highly regulated through a system of permits. Depending on the waste category, stricter regulations apply. As such, there is an incentive to detect incorrect waste category labeling not only as a form of fraud but also as a result of human error. In this study, we propose a method to automatically find these wrongly categorized permits. In a dataset with millions of permits, we approach this problem in a special outlier detection scenario. By proposing so-called supervised category models, we model each category in contrast to all others categories in a supervised way. Anomalies for each category stand out as the lowest scoring data points according to the fitted model. Moreover, we propose and compare several transaction aggregation methods for modeling. We present a visualization to recognize the potential permit category errors. Ultimately, our work aims to discriminate between regular and non-regular behavior within waste transports. This paves the way for novel approaches to be developed and applied, as to make the tasks of domain experts more efficient and data-driven.

Keywords: Anomaly detection · Supervised learning · Aggregation

1 Introduction

Governmental bodies commit to ensuring citizen safety, security, and trust, relying on both technology and their employees. One of such agencies is the Human Environment and Transport Inspectorate (ILT). The ILT is the supervisor of the Ministry of Infrastructure and Water Management of the Netherlands. It monitors and enforces compliance with both national and European legislation and regulations in favour of a safe and sustainable human environment, safe and sustainable transport, while maintaining confidence in housing associations. Human resources are scarce and limited to biological constraints: unlike a machine, people cannot perform their job responsibilities for an indefinite continuous period of time whilst maintaining their performance [1]. Therefore, these agencies

need to wisely choose how to allocate their limited assets to maximize impact. With the ever-growing rate of data generation, manual labour is simply not appropriate or even feasible to make valid assessments on each received waste permit. Thus, the ILT recognized the need for data science as a paramount tool for the integration of data-driven approaches into their domain of expertise.

Specifically, we focus on one of the enforcement points of this organization: waste transportation events. The Waste Shipment Regulation (WSR) comprises the legislation that companies must follow to transport waste(s) through a EU member state. The legislation requires that a company wanting to transfer waste must report it in advance. The report consists of relevant information such as the type of waste being transferred, its origin and location regarding country and place, and the companies related to the transaction, the total amount of waste. The report must be then sent to and processed by the ILT. In addition to the aforementioned report, companies must either make a deposit or obtain a bank guarantee to prove that they can bear all risks associated with each specific waste transfer. Since certain waste types have less stringent regulations than others as well as different deposit fees, some companies might intentionally mislabel their waste. By mislabeling types of waste and not abiding by the stipulated legislation, negative environmental impacts may occur such as the contamination of soil and bodies of water which, in turn, deteriorate the health and safety of the general populace. It is imperative to find these potentially mislabelings automatically. Hence we propose a novel method of anomaly detection applied to this domain with the aim of aiding the inspectors in targeting anomalous notifications.

Each permit consists of a sequence of transport events. Risks can be established per permit or per transport. Data will be processed in two distinct manners: transport and permit. For the permit outlier model we propose three transport aggregation schemes. Our anomaly detection proposal consists of creating a predictive model for each waste type that, given training data, learns to perform classification of its associated waste category. We focus on the individual decision probabilities (scores) each model yielded with regards to each data point. Generically, this approach intends to discriminate between regular and non-regular behaviour of data with regards to its category or class. Since we have the ground truth of waste category for each data point, we propose the lowest scoring points to be most probable anomalies.

The next section regards previous work performed in the fields of outlier and anomaly detection, regarding also data quality. The third section will focus on describing the data provided by the ILT. The following section details our methodology, followed by the experimental results. Lastly, we discuss and conclude our findings while additionally proposing improvements to both our approach as well as its application in the real world.

2 Related Work

Here, we discuss related work on outlier detection. An outlier is an observation within some data sample which strongly deviates from other observations such that it is possible to conjecture that it was generated under different conditions [2]. Identifying outliers has proven to be of great significance [3, 4]. Furthermore, the concept of outlier detection is intrinsically related to data quality assurance [5–7]. In this manner, detected outliers are not simply regarded isolated data points within some feature space. They represent, in fact, errors in data or mislabels as shown in Fig. 1. The following subsections depict unsupervised and supervised approaches to outlier detection, respectively.

2.1 Unsupervised Outlier Detection

Unsupervised outlier detection techniques have been a focal aspect of computational research, initially following the notion of outliers as objects for which a large proportion of the data lies beyond a fixed euclidean distance threshold [8]. Commonly, variations of this distance-based approach are grounded on the distance to the k th-nearest neighbor (k NN distance) [9], or on a collection of the distances to each k nearest neighbor (k NNs) [10]. Simply put, outliers are determined as data points having the largest k NN distances. Distance-based methods have similar problems to distance-based classifiers: all features are considered equally important and noisy features can overshadow significant ones. Moreover, it is impractical to measure distances between categorical variables.

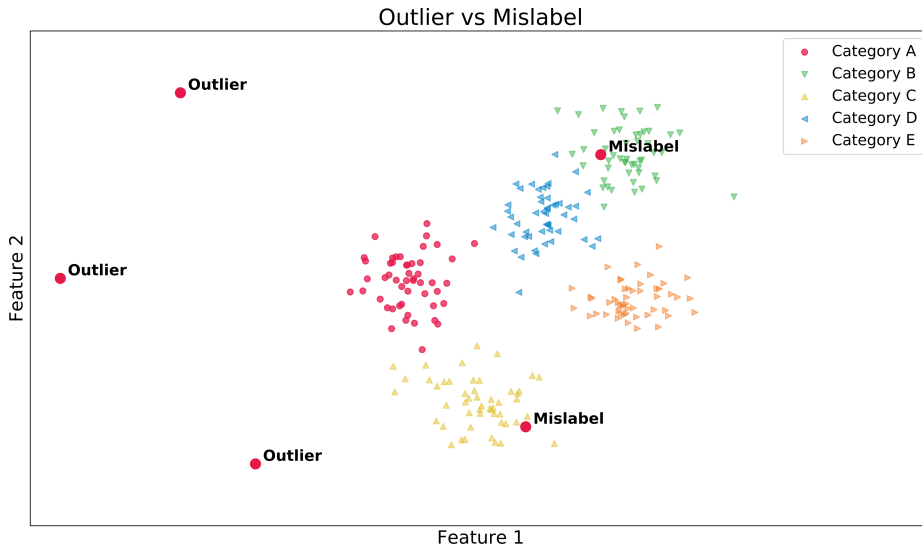


Fig. 1. Anomaly distinctions. Representation of a multi-category scenario depicting the conceptual distinction between outlier and mislabel.

2.2 Supervised Outlier Detection

Supervised approaches to outlier detection require previously established labels defining what is an inlier. When such information is available, it is possible to train a classifier data with the purpose of learning to distinguish between inliers and outliers [11]. One dominant problem that must be addressed within this special case of classification problems is class imbalance [12]. Since outliers are defined as rare instances in the data, the distribution between normal and abnormal classes is very skewed. The implication of this is that the optimization of classification accuracy is probably not meaningful, as usually misclassifying outliers is more detrimental than misclassifying inliers. Our data does not have such labels. Hence, this kind of straightforward approach would not be feasible.

3 Data and Preprocessing

In this section, we describe the data provided to us by the ILT as well as the pre-processing steps taken to make our approach possible. We start by characterizing the data generation process and its features, and later dive into its subsequent manipulation.

3.1 Data Characteristics

Prior to transporting waste materials within the EU, companies must request a licence permit. The permit request contains information about the amount of waste being transported, as well as the number of actual transports of waste, waste category, names of companies involved, etc. The ILT receives these permits requests and stores them for inspection. In this work, we utilize data collected

Table 1. Feature characterization.

Feature	Description	Example
Date	Timestamp of notification	2009-01-13
Type	Direction of a transaction	Import
Client	Name of sending company	Some Company A
Origin	Country of sending company	Australia
Border	Location of transporter border crossing	Port of Rotterdam
Request	Total tonnage requested per unique permit	150.0
Purpose	Objective of waste transport	Useful application
Tonnage	Tonnage of single waste transportation event	24.25
Processor	Name of receiving company	Some Company B
Transport	Infrastructure used for transportation	Road
Waste code	European Waste List code (1–20; target)	16
Destination	Country of receiving company	Netherlands
Licence code	Unique identifier of a licence permit	AA000001
Client location	Place of sending company within its country	Langenhagen
Processor location	Place of receiving company within its country	Dordrecht

Table 2. Waste code characterization.

Code	Waste category
01	Exploration, mining, quarrying, physical and chemical treatment of minerals
02	Agriculture, horticulture, aquaculture, forestry, hunting, fishing, food
03	Wood processing and production of panels, furniture, pulp, paper, cardboard
04	Leather, fur and textile industries
05	Petroleum refining, natural gas purification and pyrolytic treatment of coal
06	Inorganic chemical processes
07	Organic chemical processes
08	Coatings (paint, varnish, vitreous enamel), adhesives, sealants, printing inks
09	Photographic industry
10	Thermal processes
11	Chemical surface treatment, coating of metals and other materials
12	Shaping, physical and mechanical surface treatment of metals and plastics
13	Oil wastes and wastes of liquid fuels (except edible oils, 05 and 12)
14	Organic solvents, refrigerants and propellants (except 07 and 08)
15	Waste packaging
16	Unspecified (not otherwise specified in the list)
17	Construction and demolition
18	Human or animal health care and/or related research
19	Waste management facilities
20	Municipal wastes including separately collected fractions

by the ILT between the years of 2009 and 2015. Within such data, each row represents an individual waste movement. In every entry, a waste code reflecting its category is present. Moreover, several waste movements can be linked to a unique licence permit identifier. This collection of rows represents a permit request. An exhaustive list of variable names and descriptions can be found in Table 1. Waste category definitions are shown in Table 2.

3.2 Preprocessing

These requests are human-generated, which leads to issues related to the quality of data and its usability. The major concern is related to entity resolution, where different identifiers exist related to the same entity. As an example, let us consider the hypothetical *Client* value "Some Company A". Within the raw dataset, different spellings of this company name can be witnessed such as "some_company_A", "som_comp_a", or "some_company_A.inc.". To appropriately consider different values as being representative of the same real world concept (*i.e.*, a name of a country, place, or entity), an equality metric must be deployed to compare them. Typically, the Levenshtein method is used [15]. By setting a threshold on how similar a pair of strings should be as to be considered the same, this metric provides a semi-automated approach to deal with entity resolution. Every subset of a pair of values at least 80% similar is highlighted for merging. This threshold was selected based on the number of matches presented. Since the scale of our dataset allowed for it, we could manually validate

the suggested entity pairs. Using lower threshold values proved to be impractical as too many matches that did not represent the same entity were presented (even with the scale of our dataset). The initial entity list of size 2379 was reduced to 2056 (13.58% decrease). Because most of our features are categorical, dummy variables were produced in such cases to make use of them in our task.

4 Approach

Here we describe our methodology. The first two subsections relate directly to our aggregation approach with regards to transports and permits, and the strategies used to do so. Following, we characterize our conceptual approach as well as the steps required to implement it.

4.1 Aggregation

An important aspect of data processing that is the fulcrum of our experimental setup is data aggregation. A licence permit is a collection of transports (rows in the original dataset): a company issuing such a licence must indicate which individual waste transports will occur. Data is, therefore, generated in bulks of rows pertaining to the same licence identifier. Given this real-world scenario, it makes sense to handle data not only as individual transport level but also at the collective permit level. In this manner, each row refers to a licence permit. Therefore, we deal with two datasets: transports and permits.

With the exclusion of *Date* and *Tonnage*, all other attributes are the same for rows referring to a licence permit. In other words, all original rows with the licence identifier "AA000001" have the same company names, locations, etc. Accordingly, to collapse towards a permit, only *Date* and *Tonnage* need to be aggregated. With regards to aggregation, a new column *Num transports* is generated, indicating the total number of transports. Furthermore, the timespan

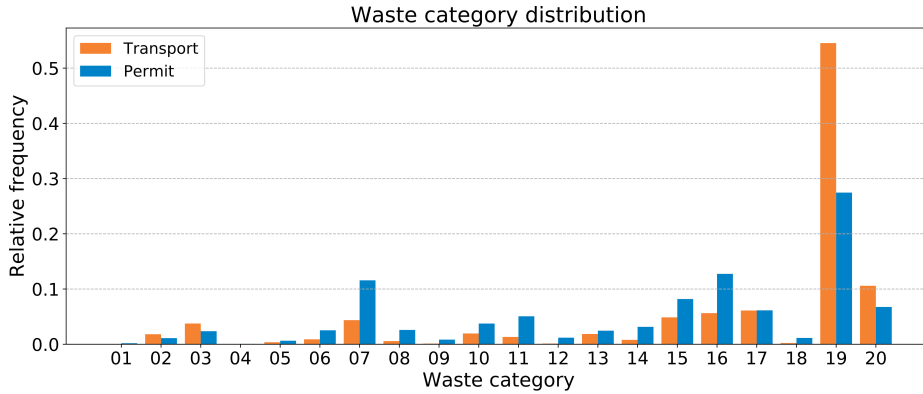


Fig. 2. Relative frequency of each waste category per dataset.

of the licence permit is registered as the difference in days between the first and last dates of the permit — *Duration*.

Aggregating transports into permits alters the relative frequency of each waste category from the transport dataset to the permit dataset as seen in Fig. 2. By viewing the aforementioned figure, strong class imbalance is clearly visible. Making use of aggregation techniques, such imbalance is attenuated; for example, the overall relative frequency of waste category 19 diminished from over 0.5 to under 0.3; except for classes 02, 03, 04, 19, and 20, all other category frequency values rose).

4.2 Binning

We want to capture as much information as possible when aggregating transports into permits. Common approaches to this issue involve computing the mean and variance of values being collapsed, as well as recording their minimum and maximum values [16]. In this manner, however, the global distribution of the feature being aggregated is not encoded and information is lost.

For *Tonnage*, a new column is created representing the sum of tonnage values of a permit; additionally, several new columns are generated representing different tonnage intervals. These intervals (or bins) are computed in 3 distinct methods: linear spacing, logarithmic spacing, and equal-frequency spacing. For every method, we first take all *Tonnage* values within the transport dataset. We create cut-off points depending on the method being applied. Linear and logarithmic methods create linear and logarithmic cut-off points, respectively. Equal-frequency binning creates cut-off points such that all bins have equal proportion. Cut-off points were calculated using the whole transports data base. For a set of intervals, relative frequencies are attributed according to the original values being aggregated into a permit summing up to one. Hence, we have successfully encoded the distribution of this feature.

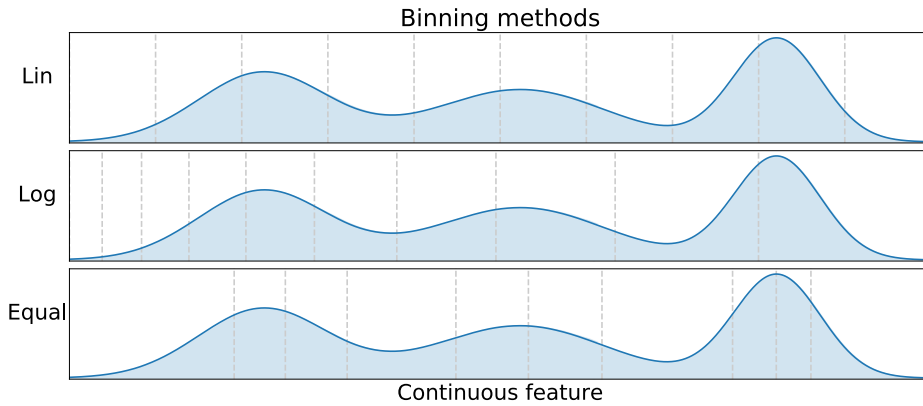


Fig. 3. Binning methods concept. Every method has the same number of bins. Cut-off points are represented by dashed lines.

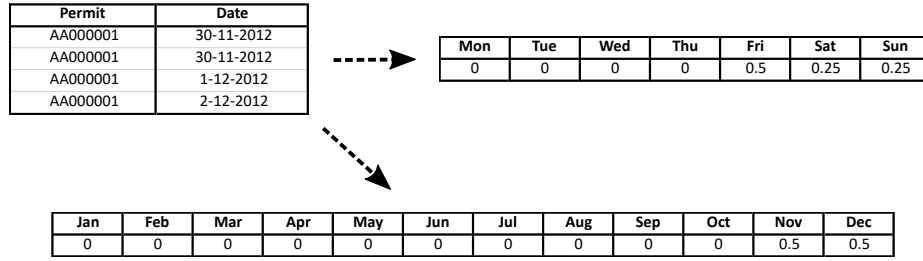


Fig. 4. Binning concept of *Date* feature. In this case, a permit is a collection of four transports: two transports in November (both on a Friday) and two in December (one on a Saturday and the other on a Sunday).

To aggregate *Date* values, we propose a binning approach where months and days of the week when transports occurred are encoded. In this sense, we will have a relative frequency distribution of the permit-related transports. The sum of values attributed to months must be one, as with days of the week. An example of this approach can be regarded in Fig. 4 where a set of hypothetical transports are aggregated into a single permit.

4.3 Supervised Category Models

As our work strives on the notion of an anomaly as an outlying data point specifically with respect to category comparison (mislabeling), an unsupervised approach would not satisfy our needs. In other words, our goal is to find mislabels within data, not outliers in a broad sense. Recalling Fig. 1, an unsupervised outlier detection model would most likely output all annotated ("Outlier" and "Mislabel) points as outliers with respect to category A. We aim to discriminate, however, only the entries marked as "Mislabel": data instances with feature values that are more similar to some category other than the one their label suggests. Hence, we are not simply interested in distance-based output.

We engage this problem in a supervised manner where we model waste categories in a one-versus-rest approach and assess how likely a data point is to pertain to its category in relation to all others. In this sense, the lowest scoring permit or transport with regards to its labeled waste category is regarded as most probable anomaly related to data consistency (mislabel) (Fig. 5). Thus, we define an category-specific anomaly/mislabel as an outlying data point with regards to other classes. The following step is to regard the scores attributed to data points belonging to the category being modeled (Fig. 6). The lowest predictive values are regarded as most probable anomalies with regards to waste category and are marked as possible mislabels for inspection by domain experts. Note that these mislabels may be indicative of not only poorly categorized waste entries, but also input errors with respect to other variables.

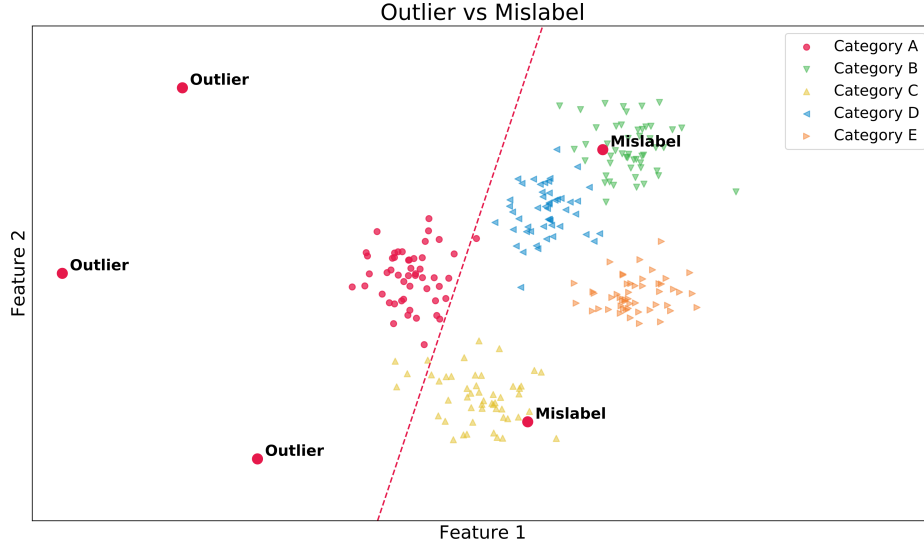


Fig. 5. Model fitting with regards to category A in a one-versus-rest approach. The dashed line represents the decision boundary of this hypothetical scenario.

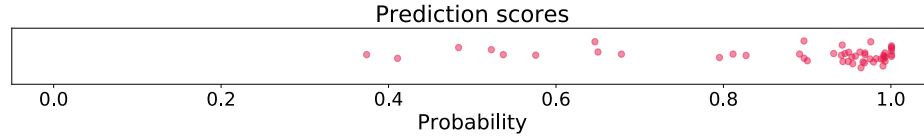


Fig. 6. Prediction scores of hypothetical data points belonging to category A after model fitting. The two leftmost points correspond to the previously annotated mislabels. Gaussian noise was added to the vertical axis for display purposes.

4.4 Implementation

Both transport and permit datasets are treated identically in terms of our classification approach. However, the permits dataset is unfolded into three subsets: one for each *Tonnage* binning methods. A total of four datasets were processed: transports, permits with linear binning, permits with logarithmic binning, and permits with equal-frequency binning. We implemented a scalable tree boosting classifier approach (XGBoost [17]) due to its competitive performance in various application domains [18–20] and, specifically in our case, its capability of handling mixed data appropriately [21]. The default XGBClassifier hyperparameter settings (https://xgboost.readthedocs.io/en/latest/python/python_api.html) were used.

For every dataset, each waste category is fitted using the entire set of rows (no cross validation). The resulting fitted scores of data points with regards to its waste category are stored. We additionally performed 10-fold stratified cross-validation to assess classifier performance in every dataset, for each waste cat-

egory. Scores are then converted to estimated probabilities (logistic regression) and for each waste category-dataset pair. Finally, performance metrics are computed by using all 10-folds simultaneously; *i.e.*, the 10-fold outputs are merged into one list prior to computation. This is done to deal with the issue of low frequency categories: for example, class *01* ($n = 22$) would produce unreliable performance sampling if divided into 10 subsets of approximately $n = 2$ each.

Given that our notion of anomaly is rooted on the concept of class comparison, better performance directly correlates to a higher confidence regarding anomaly ranking. That is, should a given model perform well, indicating it does a good job at distinguishing a particular class, then the resulting predictions with lowest output probability can be considered as most probable anomalies of that class. Should a classifier perform poorly then the resulting output scores become less reliable, as the classifier is not able to properly model its class. The performance measures used were the area under the ROC curve [22], and average precision due to class imbalance [23]. Also due to class imbalance, discrepancies between ROC AUC and average precision values are to be expected. For the three distinct aggregation methods, the datasets with highest performance per waste category were selected.

5 Results

In our experimental setup, we have four datasets. One dataset being related to transports as rows, and the remaining three relating to permits with different binning approaches (linear, logarithmic, and equal-frequency). For each dataset, we computed performance metrics per waste category as detailed in the previous section. Our aim is to regard these metrics as indicators of performance for the different aggregation methods, while also comparing them to the non-aggregated dataset. Ultimately, the classification scores per waste category are to be regarded as indicators of abnormality within the data.

We present firstly classifier performances output from the one-versus-rest 10-fold cross-validations (Table 3 and Table 4). Performance is quantified in terms of average precision and area under the ROC curve, per transport and permit datasets. Recall that values were obtained with one single run of performance metric computation, where each set of 10 folds was used merged. Hence, no values of category-specific standard deviation are present.

Both logarithmic and equal-frequency binning methods exhibited similar results (difference in metrics ≤ 0.01 between the two methods). Linear binning systematically underperformed in comparison with a difference ≥ 0.1 for all metrics when compared to either logarithmic or equal-frequency binning methods. Waste category *04* had insufficient occurrences for cross-validation (eight in total) in the permit dataset and is not displayed in the performance results for that reason. We present also the prediction output both datasets (Fig. 7 and Fig. 8) to illustrate the differences between waste category score distributions.

Table 3. Average precision per class model with regards to dataset. Overall mean and standard deviation values of average precision across all waste categories are 0.895 and 0.135, and 0.679 and 0.207, for transport and permit datasets, respectively.

		Average precision																			
Transport		0,426	0,995	0,999	0,737	0,843	0,924	0,886	0,815	0,976	0,982	0,979	0,914	0,977	0,719	0,954	0,951	0,991	0,869	0,994	0,958
Permit		0,339	0,768	0,914	–	0,68	0,568	0,777	0,494	0,843	0,754	0,849	0,756	0,673	0,578	0,769	0,721	0,841	0,684	0,891	0,688
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
		Waste category																			

Table 4. Area under the ROC curve per class model with regards to dataset. Overall mean and standard deviation values of ROC AUC across all waste categories are 0.996 and 0.003, and 0.952 and 0.021, for transport and permit datasets, respectively.

		Area under the ROC curve																			
Transport		0,998	0,999	0,999	0,999	0,999	0,996	0,99	0,994	0,998	0,998	0,995	0,998	0,998	0,99	0,995	0,992	0,999	0,998	0,995	0,994
Permit		0,939	0,968	0,988	–	0,938	0,927	0,956	0,947	0,979	0,951	0,985	0,944	0,959	0,932	0,953	0,926	0,968	0,987	0,941	0,929
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
		Waste category																			

6 Conclusions and Future Work

Our main goal in this work was to devise an automated detection system for mislabels in waste data. For that purpose, we presented a supervised waste category model where mislabels are considered as lowest scoring entries within a waste category in relation to all others. We proposed also to use three aggregation strategies for (numerical) tonnage values of the individual transport for a waste permit and assess their applicability. Considering the different aggregation strategies we followed and their subsequent evaluation, we were able to assess that for our data a linear-based approach was not appropriate. Additionally, we may conclude that logarithmic and equal-frequency binnings strategies provide equally viable options for aggregation.

The discrepancy between ROC AUC and average precision values visible in cases like category *01* is due to low instance frequency with respect to other classes (class imbalance) as stipulated previously within this work. Regarding highest performance categories, in particular class *03*, an initial assessment of its lowest transport dataset data point revealed an input error within *Tonnage*. Concretely, its presented value for that variable was off by three orders of magnitude when compared to other transport rows pertaining to the same permit identifier. This is likely because during data insertion a comma was mistaken for a dot.

Having domain experts investigate all waste transportation events is impossible: the number of notifications generated per unit of time is too large given the amount of time and human resources available. We must cooperate with domain experts to check whether the anomalies predicted by our model are in-

deed associated to non-compliant transportation events. However, we do pave the way for further data-driven approaches to be developed and applied within this particular domain. This will enable us to improve their work methodology and efficiency in a foreseeable future, while improving on our own approach with domain expert feedback. To follow in this direction, for example, a real-time system could be developed to assist governmental agencies in visualizing and detecting outliers.

References

1. Wright, T., Bonett, D.: The contribution of burnout to work performance. *Journal of Organizational Behavior* **18**(5), 491–499 (1997)
2. Hawkins, D.: Identification of outliers. 1st edn. Chapman and Hall, London (1980)
3. Aggarwal, C.: Outlier analysis. 1st edn. Springer-Verlag, New York (2013)
4. Janssens, J.: Outlier selection and one-class classification. PhD Dissertation, Tilburg (2013)
5. Zhang, Y., Meratnia, N., Havinga, P.: Ensuring high sensor data quality through use of online outlier detection techniques. *International Journal of Sensor Networks* **7**(3), 141–151 (2010)
6. Kauffmann, A., Huber, W.: Microarray data quality control improves the detection of differentially expressed genes. *Genomics* **95**(3), 138–142 (2010)
7. Kandel, S., Parikh, R., Paepcke, A., Hellerstein, J., Heer, J.: Profiler: integrated statistical analysis and visualization for data quality assessment. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp 547–554. ACM, New York (2012)
8. Knorr, E., Ng, R.: Algorithms for mining distance-based outliers in large datasets. In: *Proceedings of the 24th International Conference on Very Large Data Bases*, pp. 392–403. Morgan Kaufmann Publishers Inc., San Francisco (1998)
9. Ramaswamy, S., Rastogi, R., Shim, K.: Efficient algorithms for mining outliers from large data sets. In: *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp 427–438. ACM, New York (2000)
10. Angiulli, F., Pizzuti, C.: Outlier mining in large high-dimensional data sets. *IEEE Transactions on Knowledge and Data Engineering* **17**(2), 203–215 (2005)
11. Chow, M., Sharpe, R., Hung, J.: On the application and design of artificial neural networks for motor fault detection. *IEEE Transactions on Industrial Electronics* **40**(2), 189–196 (1993)
12. Krawczyk, B.: Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence* **5**(4), 221–232 (2016)
13. Bahnsen, A., Aouada, D., Stojanovic, A., Ottersten, B.: Feature engineering strategies for credit card fraud detection. *Expert Systems with Applications* **51**, 134–142 (2016)
14. Whitrow, C., Hand, D., Juszczak, P., Weston, D., Adams, N.: Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery* **18**(1), 30–55 (2009)
15. Levenshtein, V.: Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady* **10**(8), 707–710 (1966)
16. Irawan, C., Salhi, S.: Aggregation and non aggregation techniques for large facility location problems—a survey. *Yugoslav Journal of Operations Research* **25**(3), 313–341 (2015)

17. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 785–794. ACM, New York (2016)
18. Friedman, J.: Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* **29**(5), 1189–1232 (2001)
19. Torlay, L., Perrone-Bertolotti, M., Thomas, E., Baciú, M.: Machine learning-XGBoost analysis of language networks to classify patients with epilepsy. *Brain Informatics* **4**(3), 159–169 (2017)
20. Nielsen, D.: Tree boosting with XGBoost—why does XGBoost win “every” machine learning competition? Master’s thesis, NTNU (2016)
21. Hastie, T., Tibshirani, R., Friedman, J.: *The elements of statistical learning*. 2nd edn. Springer-Verlag, New York (2008)
22. Fawcett, T.: An introduction to ROC analysis. *Pattern Recognition Letters* **27**(8), 861–874 (2006)
23. Saito, T., Rehmsmeier, M.: The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE* **10**(3), e0118432 (2015)

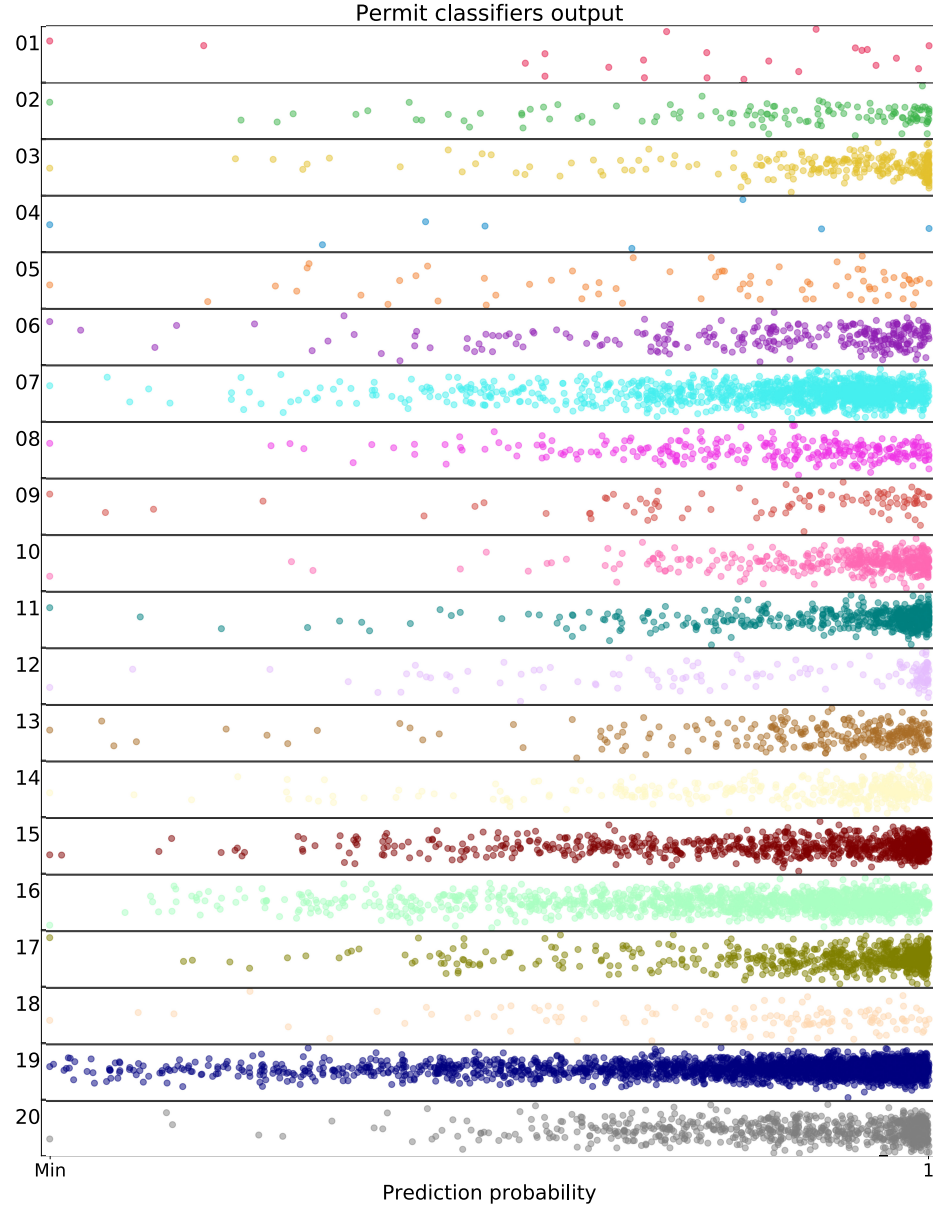


Fig. 7. Classifier output on permit dataset. Each row represents the classifier output for each waste category. Waste categories are represented by their respective code. Gaussian noise was added to the vertical axis for display purposes. The horizontal axis represents the output score probabilities and has been scaled so that the minimum value of each output is leftmost and the maximum to the far right.

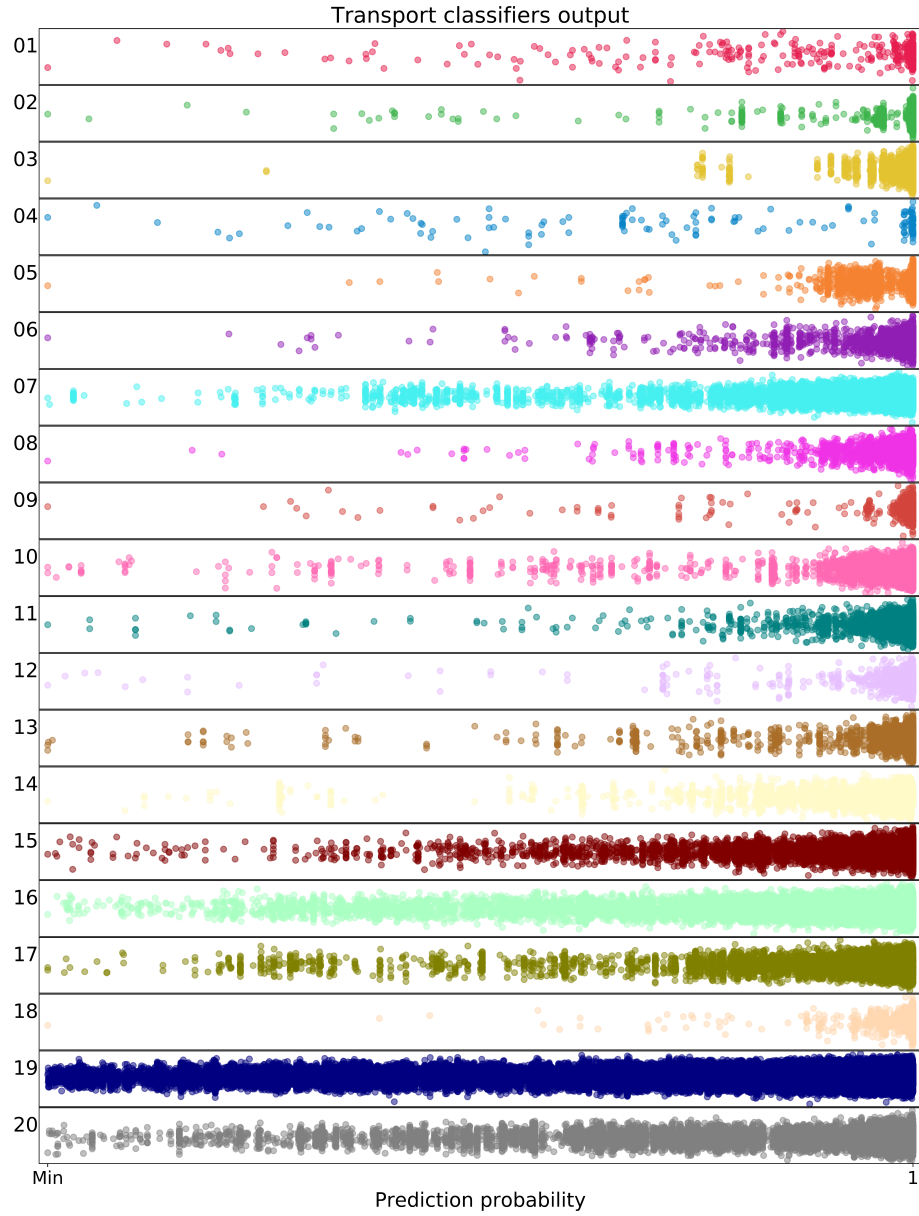


Fig. 8. Classifier output on transport dataset. Each row represents the classifier output for each waste category. Waste categories are represented by their respective code. Gaussian noise was added to the vertical axis for display purposes. The horizontal axis represents the output score probabilities and has been scaled so that the minimum value of each output is leftmost and the maximum to the far right.