Optimize Capacity for a Uniform Waste Transportation Collection

José Tiago Silva[1], André Filipe Oliveira[1], Ana Lúcia Martins[2(✉)], and João Carlos Ferreira[1,3]

[1] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Lisbon, Portugal
{jtbpc,afmao,jcafa}@iscte-iul.pt
[2] Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL), Lisbon, Portugal
almartins@iscte-iul.pt
[3] INOV INESC Inovação—Instituto de Novas Tecnologias, 1000-029 Lisbon, Portugal

Abstract. Transportation-related costs are responsible for a large portion of the waste collection process. In the past several optimization approaches in routing having been the proposal with a diversity of algorithm. In this work we propose a novel approach where we analyze waste deposition volume and try to identify patterns for a deterministic and uniform waste collection. Instead of routing optimization we propose a capacity determination based on location, year period, special events and weather conditions. An IoT sensor transmitted volume every time the wasted door is open and provide real-time value.

Keywords: Frequency-capacity · Logistics · Transportation · Waste collection · IoT

1 Introduction

In smart cities, the use of technology is common to optimize several services provided by the city council. One of the areas where technology can be used is in waste collection. By adding sensors to containers with the ability to measure the volume of waste in it, each time the container is opened, it's possible to know, in real time, the volume of waste in every container of the city. In Portugal, this is already used in cities such as Castelo Branco, but all the data generated by the sensors are typically used for routing optimization only.

Problems like frequency-capacity optimization with a fixed frequency of waste collection or the correlation of waste data with other datasets are not typically addressed. The frequency-capacity optimization problem consists in, given a frequency of waste collection (like twice for a week), what is the best number of containers by geographic area so that there are no filled containers.

This work aims to explore the data generated by the sensors and the correlation of that data with other data sets, according with, events or atmospheric conditions. It also aims the design and implementation of an algorithm-based analysis to solve the problem of container frequency-capacity optimization by location. To do this, we

A. L. Martins et al. (Eds.): INTSYS 2019, LNICST 310, pp. 108–128, 2020.
https://doi.org/10.1007/978-3-030-38822-5_8

analyze real data on the volume of containers over time in Portuguese cities between in the years of 2017 and 2018 with 18 thousand registers.

The problem of calculating the required capacity of containers by geographic location, fixing the frequency of waste collection, has not yet been addressed in the literature, which increases the interest of the topic addressed in this article. The results obtained can be used to save resources and costs to the city councils that decide to implement the algorithms under study. The correlation between waste volume data sets can also provide interesting information about the habits of the citizens.

2 Literature Review

Most of what has been studied about the waste collection are focused on routing problems. It's possible to associate the waste collection routing problem with the generic Traveling Salesman Problem (TSP) or Vehicle Routing Problem (VRP) [1]. The TSP consists in, given a set of n cities and the distance between them, and the best path for a Salesman to visit all the cities once and only once and return to the initial city. In the VRP, instead of one salesman or vehicle, we have m vehicles to visit n cities. In waste collection optimization, the containers represent the cities and the garbage trucks represent the vehicles. To limit the waste collection schedule, it can be added time windows restrictions to this problem [2]. Despite their simple statement, both these problems are too complex to solve obtaining the optimal solution when the number of containers is large [3], so it's typical to see heuristic approaches to obtain good solutions in less time [4].

Several articles study this problem, proposing algorithms for the calculation of good routes using optimization and/or machine learning. In [5], a mathematical formulation of the problem is presented and several papers in the literature are classified by the type of algorithms proposed. In [6], a genetic algorithm is presented for the identification of optimal routes for Municipal Solid Waste collection, supported by a geographic information system. Good solutions were achieved but for a small and simplified waste collection routing problem. In [7], the proposed algorithms differ from the previous ones in the literature because they are dynamic algorithms and at the same time robust, being prepared for the recalculation of the routes in the event of any failure or of a collection truck reaching the limit of capacity.

Some papers focus on optimizing time and costs of waste collection in particular cities, like Xangai (Pudong area) [8] or Allahabad [9], proposing municipal solid waste management systems suitable for those particular places. [10] summarizes similar papers for the United Kingdom. Focused on the logistics involved in waste collection in several European cities, [11] carries out a detailed study on how to manage waste collection and what standards are imposed by the European Union. This study provides a set of current and interesting information about the problem as well as what is expected in the resolution of the problem.

More focused on cloud technologies, the article [12] presents a whole system for the collection of waste in smart cities, proposing different solutions for different stakeholders in a city. To collect data, the authors use not only the sensors but also the surveillance system of a city and it addresses several possible problems in the collection of waste in the containers. Similarly, [13] used sensors that can read, collect and transmit trash volume and used this data to calculate new routes in real time, guaranteeing that when trashcans become full, they are collected on the same day. However, by doing that, they increased the waste collection frequency too much, incrementing the daily collection cost between 13–25%.

In [14], the authors focused on forecasting quantity and variance of solid waste and its correlation with other sets of data, like residential population, consumer index and season, in Shanghai. The work [15] proposes a new architecture for the dynamic scheduling of waste collection considering the capacity of the same using sensors for their measurement. This is one of the most complete articles in the use of measurements of capacities of the containers for the calculation of the frequency of garbage collection and the calculation of routes in real time taking these data into account.

Even though there's many articles dedicated to routing optimization, it can't be found in literature a study about the frequency-capacity optimization with a fixed frequency of waste collection. This can be modelled by a generic optimization problem where we want to have the minimum number of containers needed by the geographic area that guarantees enough capacity (or maximize that capacity) with the constraints of the collection frequency. It can also be viewed as a multi-objective optimization problem where we want to minimize the total number of containers and maximize the capacity by geographic area.

3 Approach Developed

The data collected from a Lora volume sensor. Every time the door is open a volume measurement is sent to a management system. We use data from a Portuguese company Evox (www.evox.pt). The central system provides visual information about the status of every monitor waste container. Based on a pre-defined filled volume a collection route is defined, like the example provided in Fig. 1.

Fig. 1. Current solution of the waste management system, with route optimization based on capacity available.

Our approach is the data analysis to identify deposition patterns for years periods (e.g. summer, winter), correlate with special events and weathers conditions in order to determine what container capacity should be installed, for a uniform week garbage collection. To study this problem of capacity optimization given a fixed frequency, we start by analyzing sets of data of containers volume in time. Each container has a sensor that measures the volume of waste in it, every time the container is opened. The data from each container consists on the following elements: *container Id*, *description*, *container type*, *waste type*, *geographic localization*, *address*, *localization zone* and sets of reading *date and time* and respective *volume* filled in percentage. Table 1 shows an example of those elements, representing the core data of the container and data about the volume reading.

Table 1. Data set examples.

Field	Example
Container id	15415
Description	Container 611
Container type	Four weal with 1000 L
Waste type	Solid urban waste
Geo localization	39.826069/−7.493849
Address	R. do Arco do Bispo 21
Localization zone	Castle zone
Reading date and time	08/06/2018 12:04; 08/06/2018 17:21; …
Volume	59%; 83%; …

This data must be cleaned and organized in appropriate structures to begin their mining. To do so, we decided to work with the Python, because of its simplicity to manipulate datasets.

We added also weather information from the National Centers for Environmental Information (NCEI) using the information on temperature and rain that we divided into pre-defined classes. For events, we've created a crawler to find local news from 2017 to 2018 and identify the type of occurrence. Hence, with this new evidences, new classes have been added: *precipitation [mm]*; *air-temperature [Celsius]*; *type of day*; *events*, that we collected from local news, such concerts, parties, public holidays and others.

The dataset containing the information from all classes provides a big portion of the information we intend to use in the study of the capacity-frequency problem. However, because the volume is measured each time a container is opened, these discrete data doesn't have a fixed time period between readings. One container can be opened ten times in a day, while others might not be opened in that space day.

To deal with this, we created a function that generates another dataset in which the volume data frame is defined with a fixed time period of every x hours (8 h, 16 h, or even 1 day). Each line of the data frame has, for each container, information about the last measured volume and the mean and median measure of volume in that time period. This can also be viewed as a continuous dataset in which the volume of a container on a *datetime* is the last measure or the average volume in the time period containing that *datetime*. We expect with this dataset to easily get information about the average volume growth by container or zone and to have two different approaches in this study.

With the datasets defined, we present in the next section a detailed study of the information on those datasets and a visualization of the data.

4 Data Visualization

The main dataset is composed of almost eighteen thousand rows, and each row accords to a waste volume measure, a date and time, and an id of the corresponding container. In total, there are eighteen waste containers, identified by a unique id, his geographic coordinates, type of container and his total capacity. There are three types of containers: the standard ones, with 800 L and 1000 L capacity and the surface containers which can also store 1000 L.

4.1 Visualization by Zone

The containers are split across the district of Castelo Branco, making up about eight streets, as shown in Fig. 2, we can visualize the number of containers that are for disposal for each street, following by their id number and capacity.

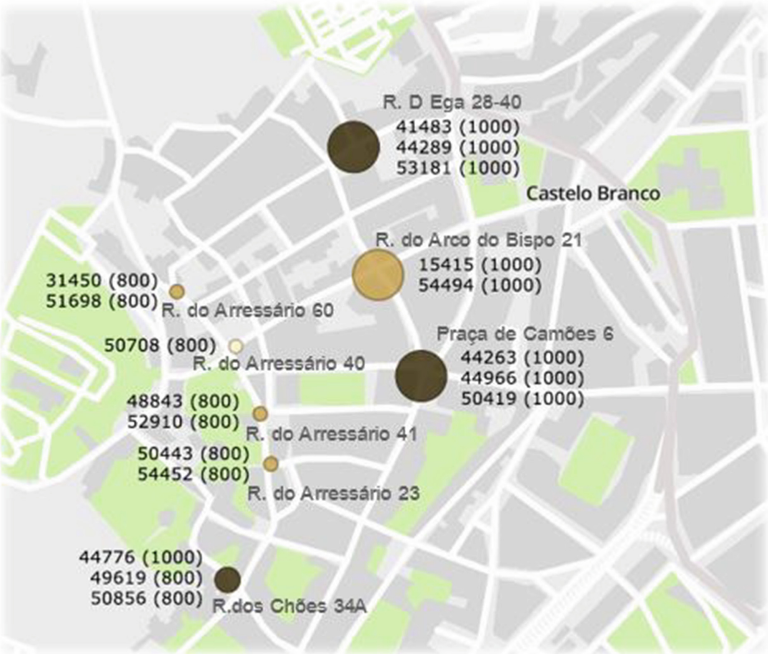

Fig. 2. Container's streets locations with a perspective view.

Here we grouped all the data by their locations mentioned before, between the dates of 08-jun-2017 to 08 jun-2018. In Fig. 3 is shown the average volume of waste inside the containers in percentage, by each street, for each month. We can see that even for an average calculation, the values seem to appear quite aleatory, however, seems to be increasing over the time. Despite the noise, we can notice that most of the volumes are between the range of 30% to 60%.

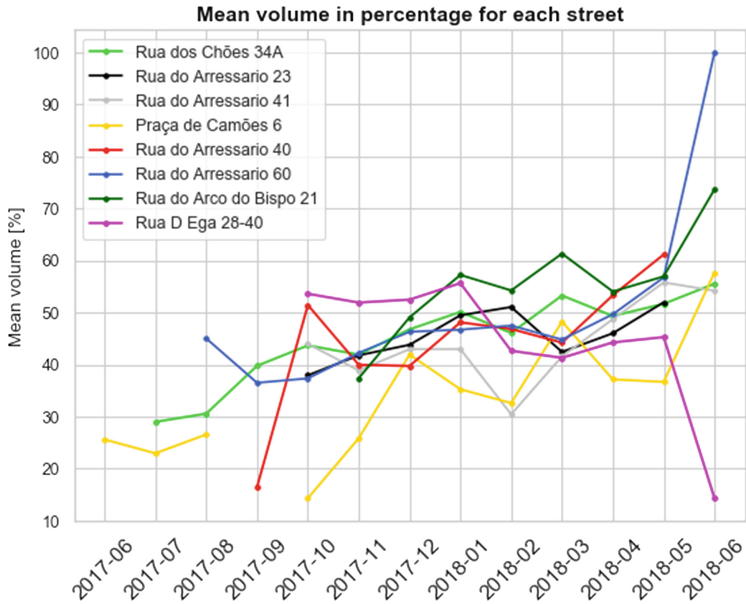

Fig. 3. Average volume of waste in the containers by street location, and their according months.

Another interesting fact is that, 90% of the times, the *volume* is below the 60%. In other words, from the full cycle of data (366 days), only in 36 days the *volume* was higher than 60% and those days mostly correspondent to the Wednesdays. This could be important later, when defining a collection day.

4.2 Deposits and Collections

With the container's locations and dates been set, the next step is to calculate the frequency of waste collection. Hence, we split the volume of waste into two types: *volume-deposits [Liters]*, which is when the volume of the containers gets filled, and *volume-collections [Liters]*, when the volume is emptied. With the class *day-of-week*, on calculated the amount of volume deposited and collected, for each day of the week. The result shown below is the average volume of liters, for every container, regardless the time of the year (Fig. 4):

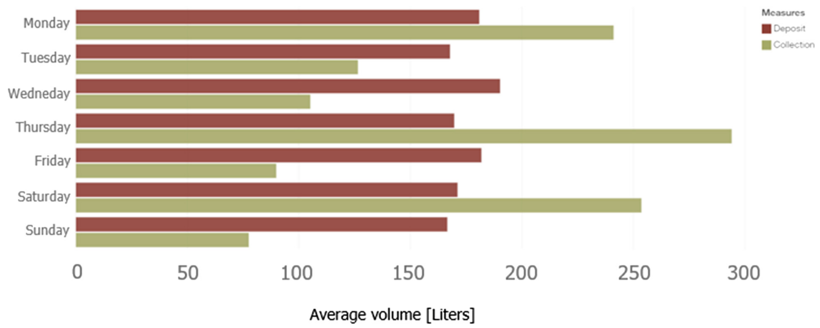

Fig. 4. Average volume of liters deposited and collected per day.

Observing at the volume collected, on can see that there are three main days of week where the waste volume is collected, at Monday, Thursday and Saturday, so the average frequency calculated is three times a week, taking into account the mean result for every container at any time of the year meaning that the frequency may vary, depending of the time of the year. We'll deduce that all the containers have the same collection day programmed, because they are very close to each other, from 40 to 80 m.

Looking at the average volume of waste deposits, we can see that there is no discrepancy between the days of week, as they vary just from 166 to 190 L, so the amount of deposits is not influenced by the day of week. However, the volume of the containers, in percentage, is always higher on Wednesdays, because is when the interval between two collections is higher.

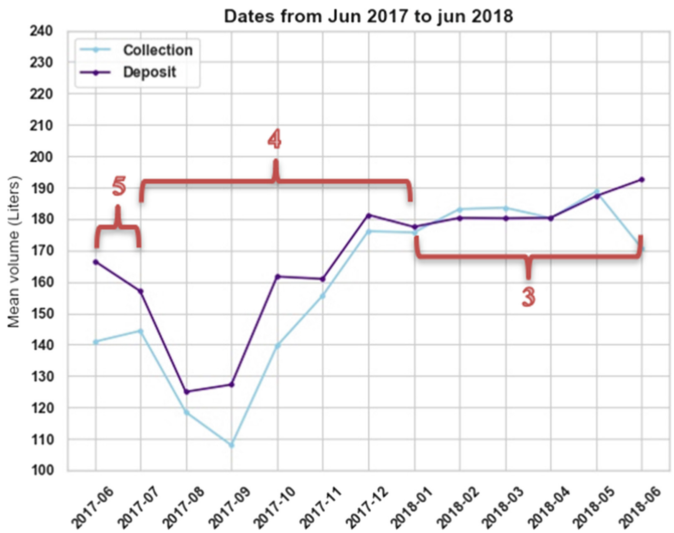

Fig. 5. Average volume of waste collection and deposit, for each month and it's waste collection frequency (Color figure online)

In Fig. 5, is registered the average volume of waste collection and deposits. By reducing the graph to the scale of one year, we can observe the volume had a higher low on august 2017, this may be due to the period correspondent to the end of seasonal time, were people come back from holydays. At that time, the deposits have been increased linearly until December 2017, and then the trend remained slightly constant from that time period (December to June). The red marked numbers shown are the occurred frequency per week, for each month. As we can notice, the frequency is dynamic, that is, it changes from month to month in order to fit the needs.

According to the frequency, observing the months from July to December 2017, the frequency was four times a week, the collections days were on Monday, Tuesday, Thursday and Saturday. From among the months between January to June 2018, the frequency has changed to a fixed amount of three (removal of Tuesday as a collection day).

4.3 Collection Analysis

Considering the amount of waste *volume* in the containers in each day and moments of waste collection, it is possible to evaluate, for each container, how well the current waste collection frequency performs. To do so, let us consider the following definitions: we consider a ***needless collection*** as the collection of waste in a container with less than 35% of volume waste and a ***critical point*** as point where the volume of a container is 100% for more than one day.

According to the collections of each container, the percentage of *needless collections* is presented in Fig. 6. On the other hand, we can see the total amount of critical points for each container in the Fig. 7.

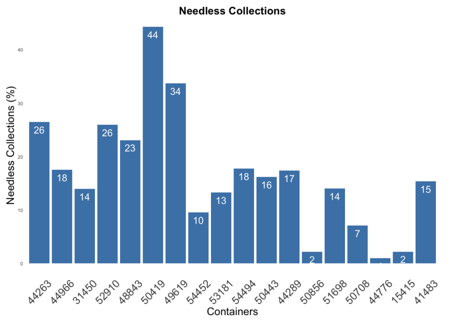

Fig. 6. Percentage of *needless collections* by container.

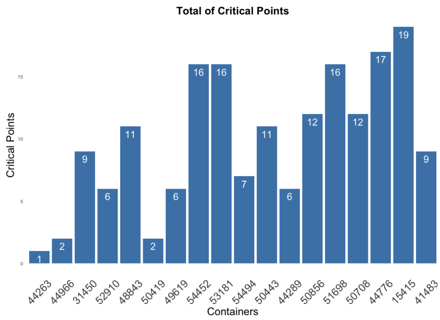

Fig. 7. Number of *critical points* for each container.

On a first analysis, we can assume that containers *44263*, *44966*, *50419* and *49619* should have less waste collection frequency, because they have high percentage of needless collections and low number of critical points. On the other hand, the containers *54452*, *53181*, *51698*, *44776* and *15415* should have a more frequent waste collection frequency, because of their high number of critical points. This data shows that the waste collections frequency and/or the capacity of the containers can be changed and improved for each container. Ideally, there would be no critical points or needless collections, but our focus will not be to minimize these points individually for each container but to consider all the containers grouped by their location and address the frequency-capacity problem.

4.4 Data Correlation

In this section, on will try to find patterns that may influence the amount of deposits. This case studies four scenarios, according to the class *type-of-day,* the database is divided into three types of day, the celebrative days but not holidays, the holidays and normal days and weekends. The class *season*, which represents the partition of the database into the different seasons, *precipitation [mm]*, which can be rainless, rain or heavy rain and *air-temperature [Celsius*], that vary from a frosty day, cold day, warm day and hot or very hot day.

Relatively to the levels of precipitation, we can notice the average waste deposits are very close to each other, showing our lower value of 165 with heavy rain, and the higher value of 180, with a normal raining day. Concerning to the air temperature, the verified values of waste deposits differ from 149 to 180 L. The amounts are also very similar, on exception of the variable *very hot day*, which is a much lower value. This may be due to the seasonal time corresponding to the summer. Comparatively to the season, on can observe the volume of deposits in the summer is significantly lower than in the rest of the seasons, as said before, there seem to appear some sort of correlation between the variables *summer* and *very hot day* and so, the values can be interpreted as the seasonal time of the year, where a set of families go out to another cities which decreases the demographic population of Castelo Branco. According to the type of day, we can relate that, in average, the amount of waste deposits is similar between the type of days, as the values are close to each other. By having a broad view of the deposit's interactions, the results vary from 160 to 190 L.

In short, on can observe that the type of day isn't really an important class, as we can see, the volume of waste deposit doesn't seem to alter from, for example a holiday to a normal day, plus, a normal day (175,6 L) presented higher volume than in a holiday (160,6 L).

4.5 Major Findings

In this section it was shown a lot of information about the dataset and a good data visualization and analysis, which will be used as leverage information for the algorithms coming in the following sections.

Regarding the class *day-of-week* on saw that that the frequency of collection is dynamic, as it may vary according to the time of the year. Also, the days of week for collection are fixed on Monday, Thursday and Saturday. Tuesday is also added when the frequency is increased to four.

The daily average volume of the containers is mostly between 30% and 60% (330 days of 366), and those few days where the volume is higher than 60%, are correspondent to Wednesdays. As the volume of deposits are, in average, about

180 L per day and the containers have capacity between 800 to 1000 L, it is possible to decrease the frequency to three times a week and in the summer to two times a week (Fig. 8).

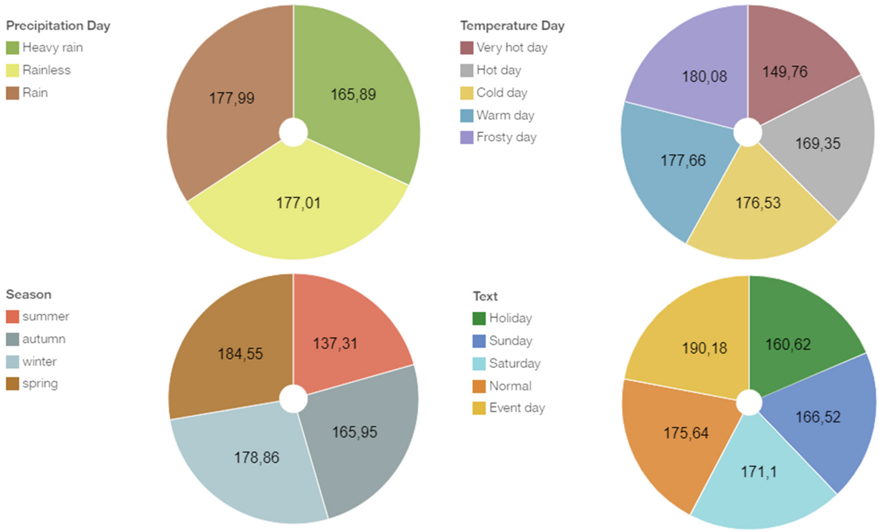

Fig. 8. Pie chart representation of the deposits by the levels of precipitation, temperature, season and type of day.

A quick analysis on the current collection shown that a large percentage of the collections are *needless collections* and some of the containers have a considerable amount of critical points, which leads to the idea that waste collection can be improved.

On data correlation, it was found that the classes *type-of-day, precipitation [mm]* and *Air-temperature [Celsius]*, haven't shown concrete results, as the variation was very low, and so, on decided to omit them, in order to delete ambiguity and posteriorly apply machine learning with the less noise as possible. Relatively to the class *season*, this indicated us that in the period accorded to summer, the volume of waste deposits decayed 175 to 130 L, which may be due to fire forests or less population density and we must take that into account.

5 Predictions

Using information such as season, events, weekday, precipitation and temperature can provide good predictions on whether a container waste must be collected or not. To do so, we used data from the main data set and several datasets with fixed time periods. In both cases we considered that a waste container must be collected if his capacity gets higher than 60%. We pretend to compare the results of the several datasets.

The classes *day-of-week*, *month*, *season*, are the main inputs and the target is *volume-filled [%]*. The inputs were used to train our machine learning model through the workflow processes illustrated in Fig. 9. Train dataset is pre-processed to align data on the same scale. Then, the processed data are fed to train the Machine Learning (ML) models where they will be hold-out and cross-validated with 80% of data. Finally, the model with chosen hyperparameters will be tested with 20% of data for testing.

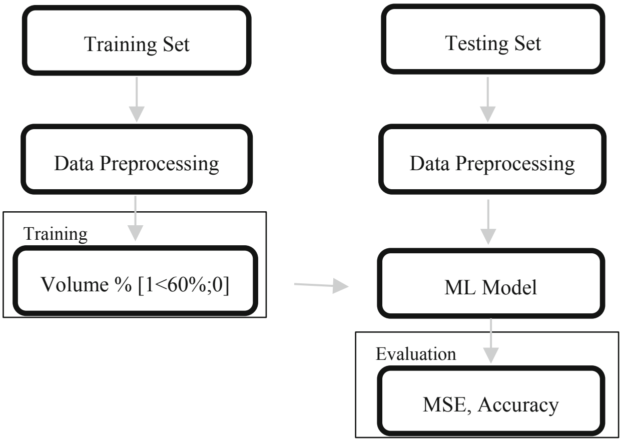

Fig. 9. Machine learning prediction process.

5.1 Data Preparation

Raw data with categorical values, such as *day-of-week, month* and *season*, are pre-processed using dummy techniques, where the number of columns is equal to the number of categories.

The target, the *volume* is what we want to predict. More specifically we want to predict if a container has to be collected. To improve the performance and match the points of interest of the article, we transferred the values, which vary from 0% to 100%, to binary data. When the volume filled is inferior to 60%, returns 1, otherwise equals to 0.

Since all the data now is composed of binary data, in exception of the class *season*, which vary from 1 to 12, we won't need to standardize nor normalize the data as all classes have the same weights.

5.2 Evaluate Algorithms

Regarding the procedures of [16], we will test the accuracy with linear and nonlinear algorithms and use 10-fold cross validation to evaluate algorithms using the **Mean Squared Error** (MSE) metric and default tuning parameters. MSE will give a gross idea of how wrong all predictions are (0 is perfect), Fig. 10.

The six algorithms selected included for the baseline of performance on this problem are:

- Linear Algorithms: Linear Regression (LR), Lasso Regression (LASSO) and Elastic Net (EN).
- Nonlinear Algorithms: Classification and Regression Trees (CART), Support Vector Regression (SVR) and k-Nearest Neighbours (KNN).

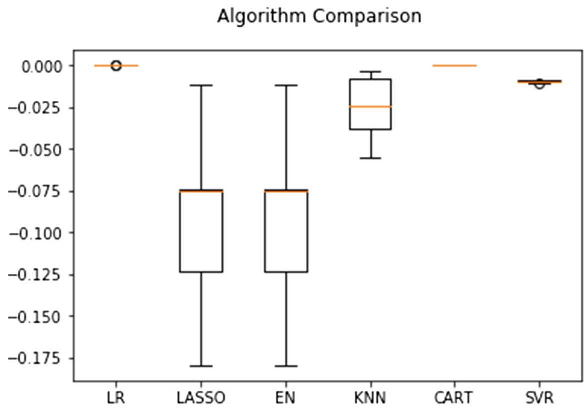

Fig. 10. Algorithm comparison (LR, LASSO, EN, KNN, CART AND SVN), using de mean square error.

On Fig. 10, represents a plot of the algorithm evaluation results and the comparison of the spread and the MSE of each model. We can see that the algorithms have a pretty good behavior, as their MSE calculated are very close to zero, in particular, LR, CART and SVR have their box and whisker plots squashed at the top of the range.

Using the metric of **accuracy** to evaluate models, which is a ratio of the number between correctly predicted and the total number of instances in percentage and using 10-fold cross-validation to estimate accuracy, we'll evaluate five different algorithms:

- Linear Algorithms: Logistic Regression (LR).
- Nonlinear algorithms: k-Nearest Neighbours (KNN), Classification and Regression Trees (CART), Gaussian Naive Bayes (NB), Support Vector Machines (SVM).

Ensuring the evaluation of each algorithm is performed using the same data splits, the results are directly comparable, in Fig. 11.

This plot shows that the accuracy of the algorithms is at least 0.92, which is a great result. This happens because of the strong correlation between the inputs and the volume data. On top of that, 90% of the time, the volume of waste is below 60%, which makes the prediction data very unbalanced and easier to predict. A study on a more balanced dataset will be made in the following subsection.

The decision tree algorithm shows, on Fig. 12 show relevant weekday on the *volume* class. In fact, the first ramification splits the dataset in Wednesday data and other weekdays data. This is consistent with the conclusions on Sect. 4, where it was shown that Wednesdays were in average the days with more waste volume.

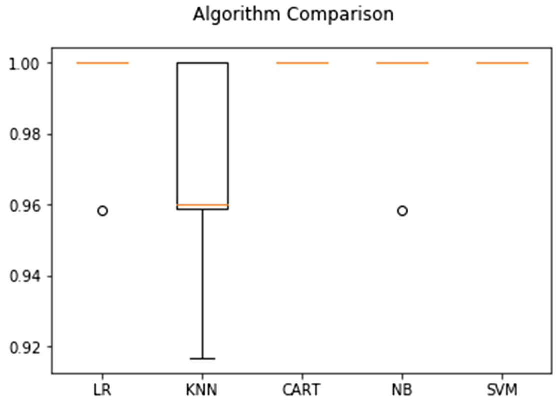

Fig. 11. Algorithm comparison (LR, KNN, CART, NB and SVN), using the accuracy score.

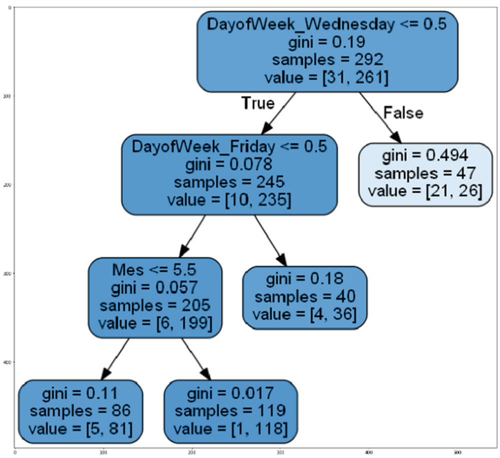

Fig. 12. Decision tree example.

5.3 Prediction with Time Periods

Considering a dataset with volume values every 8 h, for every container and using information about season, events, weekday, precipitation and temperature, we predicted if a container waste should be collected using, in this case, five algorithms: k-nearest neighbours (KNN), Latent Dirichlet Allocation (LDA), decision tree (cart) and random forest (RF). An example of the training results is presented in Fig. 13:

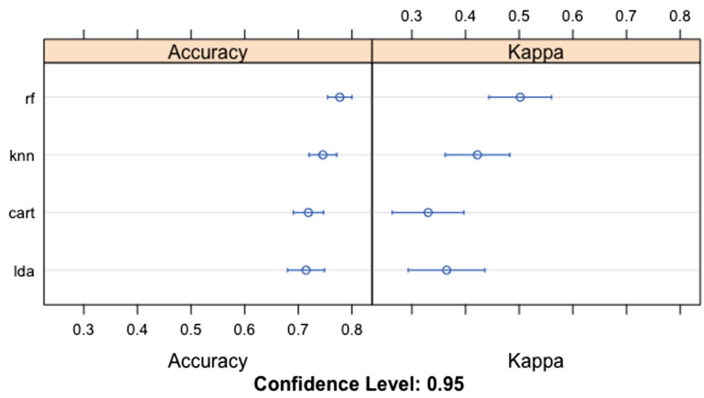

Fig. 13. Training results.

In this example, we can see that the random forest algorithms present the better results, with up to 80% of accuracy. With this, random forest was the elected algorithm for the remaining tests. Although this is not a bad accuracy result, when making predictions to compare with the validation set, the predictions accuracy doesn't go further that 75%, for most of the containers data used.

Table 2. Prediction results by time period.

Container/TP	6H	8H	12H	24H
44263	0.90	0.93	0.92	0.93
44966	0.62	0.69	0.68	0.69
31450	0.73	0.76	0.72	0.72
52910	0.76	0.79	0.73	0.76
48843	0.65	0.68	0.69	0.63
50419	0.92	0.91	0.92	0.91
49619	0.83	0.83	0.77	0.82
54452	0.58	0.58	0.57	0.61
53181	0.68	0.62	0.56	0.60
54494	0.65	0.66	0.71	0.59
50443	0.69	0.66	0.68	0.68
44289	0.64	0.69	0.71	0.73
50856	0.64	0.61	0.53	0.77
51698	0.60	0.62	0.68	0.71
50708	0.67	0.58	0.66	0.67
44776	0.58	0.67	0.63	0.75
15415	0.65	0.73	0.63	0.74
41483	0.68	0.70	0.59	0.65
Mean	0.69	0.71	0.69	0.72

All results are shown on Table 2. The mean of the accuracy obtained was around 71% for the datasets with volume values every 8 h. The same algorithm was applied for datasets with time periods of 6 h, 12 h and a day. The mean of accuracy obtained was 69% 69% and 72% respectively. We conclude with this results that classification algorithms provide better predictions using the main dataset and there is no advantage of using time periods information.

The results presented show that information like season and weekday provide good predictions on whether a container waste must be collected or not. This can be useful on creating new models of collection frequency, providing a way to study how they change as the amount of volume, not only for the dates on the datasets but also to predict how they behave in the future.

6 Capacity-Frequency Models

After the data analysis shown in section four, we concluded that the current waste collection frequency in Castelo Branco a collection between three to four times a week, most of the time on Monday, Thursday and Saturday. On the other hand, if we consider that we just need to collect a container waste if the container has more than 60% of waste volume, it was shown that more than 40% of the past collections were needless collections, meaning the collection frequency should be easily decreased.

In this section, we pretend to analyze what is the capacity needed if we reduce the waste frequency to once or twice a week and present good models to find the best day or days for waste collection. To validate these models, an analysis of the containers overload (new volumes provided by the model higher than 100%) will be made.

With the historical data from each container, it's possible to simulate what happens to the volume waste if the frequency of waste collection was fixed once a week or twice a week, for every container. For that, we fix a date (*nd*) and time (*nt*) for the new collection and, from a set with time period of one hour, we generate an entire new set of volume data, for each container. Initializing $gap = 0$, this process works like this:

1. *For every* entry of the dataset we check date (d), hour (h) and volume (v);
2. *If* $d = nd$ and $t = nt$ it's time for a new collection so we set $gap = -v$, otherwise, *if* $prev\_v - v > 10$ this was an old collection and we set $gap + = prev\_v$, otherwise gap stays the same;
3. We set the new volume for this date and time $nv = v + gap$.

The model data is the set of the new volume generated, of each container. For a model with a collection frequency of more times a week, the algorithm has several days and hours as its input.

6.1 Collection Once a Week

Considering a period dataset with time period of 2 h and a waste collection frequency of once a week (Wednesday at 10 P.M.), Fig. 14 shows an example of the new model volume, compared to the real volume with the current collection frequency:

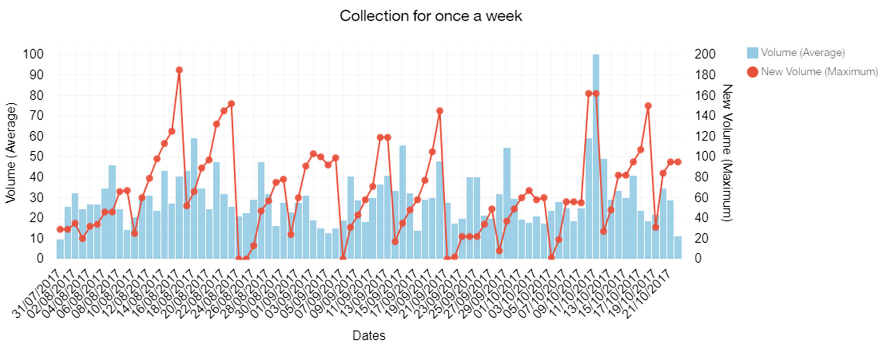

Fig. 14. First 1000 records of once week frequency for container 49619.

As we can see, a waste collection frequency of once a week is not enough for container 49619 between 06/08/2017 and 15/10/2017, with too many occurrences of waste overload.

Table 3. Once a week frequency results.

Container	Average volume	Count >100%
44263	38	46
44966	157	115
31450	108	106
52910	118	104
48843	159	124
50419	65	60
49619	80	89
54452	195	137
53181	188	139
54494	183	127
50443	152	121
44289	172	122
50856	164	153
51698	183	122
50708	140	130
44776	195	164
15415	219	109
41483	177	130
Mean	149%	117

Table 3 shows the mean of the new volume by container and the amount of waste overloads for each container. We can see by the results that a collection frequency of once a week is clearly not enough for these containers. This asks for an improvement of the container capacity or the collection frequency.

6.2 Collection Twice a Week

Considering the same container and time period, Fig. 15 shows the simulation for a twice a week frequency (Wednesday and Sunday at 10 P.M.):

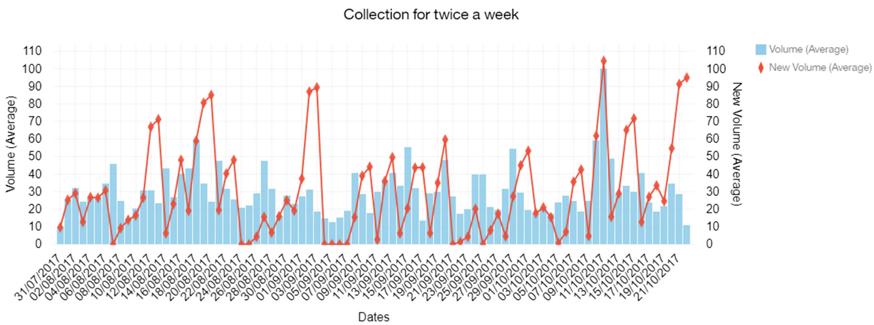

Fig. 15. First 1000 records of twice week frequency for container 49619.

We can see in this example that a waste frequency of twice a week is perfectly enough for container 49619 between 06/08/2017 and 15/10/2017, with only two occurrences of waste overload.

Table 4. Twice a week frequency results.

Container	Average volume	Count >100%
44263	19	3
44966	74	57
31450	56	59
52910	57	40
48843	77	64
50419	31	11
49619	44	34
54452	100	75
53181	90	67
54494	95	64
50443	76	55
44289	89	60
50856	81	70
51698	96	73
50708	71	49
44776	99	90
15415	117	75
41483	90	73
Mean	76%	56

Table 4 shows the mean of the new volume by container and the amount of waste overloads for each container. The results are much more reasonable, with a total average of 76% of volume.

Now, for this model we have to check the capacity by zone. The average by container or the total average do not guarantee that the current capacity is enough for this collection frequency. Grouping by streets the mean volume by month, as in Sect. 4, for these new volume sets, we have the result shown in the next figure (Fig. 16):

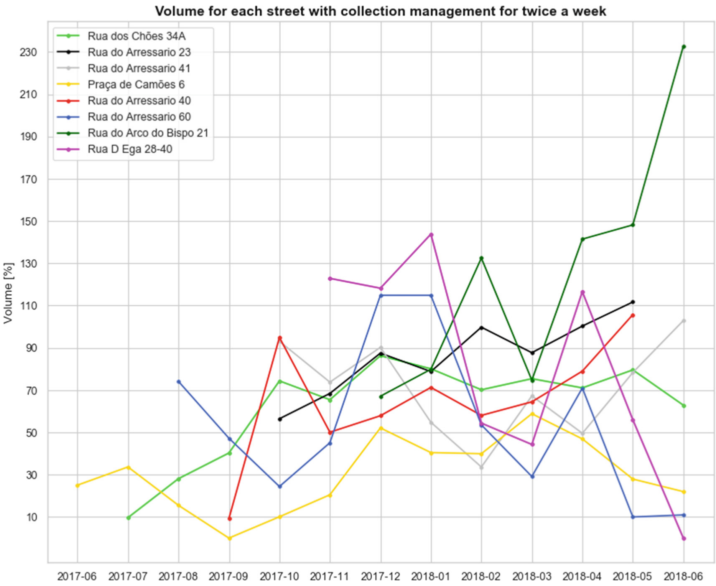

Fig. 16. Mean of volume by month grouped by streets.

This shows that the zone with more capacity problems is Rua do Arco do Bispo 34. For this zone, we have to add a container with a 1000 L capacity. For Rua D Ega and Rua do Arressário, a new container with 400 L is enough. With these improvements, we have a model with a good capacity frequency, as pretended.

6.3 Prediction on Model

To validate the new model of a collection frequency of twice a week, we applied the prediction process of Sect. 5 for each street group. The target, the *new volume* is what we want to predict, but instead of predicting if a container needs waste collection, we want to know if a container is likely to have a waste overflow. To do that, when the volume filled is inferior to 100%, returns 1, otherwise equals to 0.

The algorithm used was decision tree. The data set of the model was divided in partitions of 80% for training set and 20% for accuracy validation. The results are shown in Fig. 17.

Fig. 17. Prediction model for two times a week.

The accuracy obtained are around 75% and the predictions for this model states that the with a waste collection frequency of twice a week and the capacity changes on Rua D Ega and Rua do Arressário, represent a good solution for these containers, without waste overload.

This is a major improvement on the current collection frequency that is between three to four times a week. The addition of the containers to guarantee the capacity needed have a fixed cost, while reducing the collection frequency one or two times a week represent cost reduction every week.

7 Major Findings

In this study, it was shown (from Visualization approach) that grouping the containers by streets, the monthly average volume was always between 30% to 60% and the average volume of waste deposit was never above 20% of waste a day. On the other hand, the waste collection was, in most cases, done wrong, with a high number of *needless collections* and *critical points*. It was also found that the classes *type-of-day, precipitation [mm]* and *Air-temperature [Celsius]* had a week correlation with the volume data while *day-of-week*, *month* and *season* had a strong correlation.

Using machine learning algorithms, we predicted if a container waste has to be collected or not with a 95% accuracy, just using information on *season*, *month* and *week day*. These predictions can be used to propose more complex models where the waste collection frequency varies by season.

We propose two different models of frequency-capacity. The first proposal was a waste collection frequency of once a week. For this model, we saw that almost every container had an average waste volume over 100% which shows that a frequency of once a week is not enough for this case. The second proposal was a waste collection frequency of twice a week. This model needed a capacity adjustment for the street of Rua do Arco do Bispo, by adding a container with a 1000 L capacity and For Rua D

Ega and Rua do Arressario, a new container with 400 L. With this adjustment, the containers waste doesn't overload through all year, which makes it a successful capacity-frequency model for our containers.

8 Conclusion

In this paper we addressed the waste collection process with a different approach by studying the capacity-frequency problem.

We successfully correlated waste volume data and were able to extract information, with the variable's *year*, *season* and *weekday* which allowed us to make predictions on whether a waste container needs to be collected with a precision above 90% of accuracy.

It was possible to analyze waste deposition volume and to identify patterns for a determinist and uniform waste collection. For this case, we concluded that a uniform collection of twice a week, with small improvements on containers capacity, proved to be enough for these containers, which is a major improvement to the current collection frequency of three to four times a week.

This process is easy to implement for other sets of data because the process to generate the model's new volume data is scalable, so it's easy to apply this study for other use cases. It also allows the simulation of different waste collection frequencies in multiple periods of time.

For future work, we pretend to use the season information to propose mixed different waste collection frequencies by season and to automate the calculation of the needed capacity for a given frequency.

References

1. Nuortio, T., Kytöjoki, J., Niska, H., Bräysy, O.: Improved route planning and scheduling of waste collection and transport. Expert Syst. Appl. **30**(2), 223–232 (2006)
2. Kim, B.I., Kim, S., Sahoo, S.: Waste collection vehicle routing problem with time windows. Comput. Oper. Res. **33**(12), 3624–3642 (2006)
3. Laporte, G.: The traveling salesman problem: an overview of exact and approximate algorithms. Eur. J. Oper. Res. **59**(2), 231–247 (1992)
4. Bautista, J., Fernández, E., Pereira, J.: Solving an urban waste collection problem using ants heuristics. Comput. Oper. Res. **35**(9), 3020–3033 (2008). Part Special Issue: Bio-inspired Methods in Combinatorial Optimization
5. Ghiani, G., Lagana, D., Manni, E., Musmanno, R., Vigo, D.: Operations research in solid waste management: a survey of strategic and tactical issues. Comput. Oper. Res. **44**, 22–32 (2014)
6. Karadimas, N.V., Papatzelou, K., Loumos, V.G.: Genetic algorithms for municipal solid waste collection and routing optimization. In: Boukis, C., Pnevmatikakis, A., Polymenakos, L. (eds.) AIAI 2007. ITIFIP, vol. 247, pp. 223–231. Springer, Boston, MA (2007). https://doi.org/10.1007/978-0-387-74161-1_24

7. Anagnostopoulos, T., Zaslavsky, A., Medvedev, A.: Robust waste collection exploiting cost efficiency of IoT potentiality in smart cities. In: 2015 International Conference on Recent Advances in Internet of Things (RIoT), pp. 1–6, April 2015
8. Minghua, Z., et al.: Municipal solid waste management in Pudong new area. China. Waste Manag. **29**(3), 1227–1233 (2009)
9. Sharholy, M., Ahmad, K., Vaishya, R.C., Gupta, R.D.: Municipal solid waste characteristics and management in Allahabad. India. Waste Manag. **27**(4), 490–496 (2007)
10. Burnley, S.J.: A review of municipal solid waste composition in the United Kingdom. Waste Manag. **27**(10), 1274–1285 (2007)
11. Bing, X., Bloemhof, J.M., Ramos, T.R.P., Barbosa-Povoa, A.P., Wong, C.Y., van der Vorst, J.G.A.J.: Research challenges in municipal solid waste logistics management. Waste Manag. **48**, 584–592 (2016)
12. Medvedev, A., Fedchenkov, P., Zaslavsky, A., Anagnostopoulos, T., Khoruzhnikov, S.: Waste management as an IoT-enabled service in smart cities. In: Balandin, S., Andreev, S., Koucheryavy, Y. (eds.) ruSMART 2015. LNCS, vol. 9247, pp. 104–115. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23126-6_10
13. Gutierrez, J.M., Jensen, M., Henius, M., Riaz, T.: Smart waste collection system based on location intelligence. Procedia Comput. Sci. **61**(2015), 120–127 (2015)
14. Vicentini, F., et al.: Sensorized waste collection container for content estimation and collection optimization. Waste Manag. **29**(2009), 1467–1472 (2008)
15. Anagnostopoulos, T., Zaslavsy, A., Medvedev, A., Khoruzhnicov, S.: Top k query based dynamic scheduling for IoT-enabled smart city waste collection. In: 2015 16th IEEE International Conference on Mobile Data Management, vol. 2, pp. 50–55, June 2015
16. Brownlee, J.: Machine Learning Mastery with python (2016)