Learning Patterns in Train Position Data

Automatic Detection of Whether a Solution of the Train Unit Shunting Problem (TUSP) is a Week or a Weekend Day

Ilia Tomov[1]

Supervisor(s): Mathijs de Weerdt[1]**, Issa Hanou**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Name of the student: Ilia Tomov
Final project course: CSE3000 Research Project
Thesis committee: Mathijs de Weerdt, Issa Hanou, Jing Sun

Abstract

When not in service, trains are parked and serviced at shunting yards. The Train Unit Shunting Problem (TUSP), an NP-hard problem, encompasses the challenge of planning movements and tasks in shunting yards. A feasible shunting plan serves as a solution to the TUSP. Current automated planning tools utilized to assist human planners in this computationally heavy planning task are not able to distinguish inherent patterns in input train data, as opposed to humans. This paper aims to address this technological gap by examining whether valuable patterns could be extracted from shunting plan data, consisting of solutions to the TUSP. More specifically, it is mainly concerned with the automatic detection of whether a solution to the TUSP is a week or a weekend day. Therefore, the data is examined for the presence of several groups of patterns. Moreover, binary classification is performed on the data. The experiments conducted in this study suggest the presence of valuable patterns in the data, which could be leveraged to design specialized heuristics for automated planning models tailored to generate shunting plans for weekdays and weekends.

1 Introduction

The trains in the Netherlands serve over a million people daily. During peak hours, most of the carriages are in use to serve the passenger demand. However, at night and off-peak hours, the surplus of trains needs to be parked and serviced (maintenance, cleaning) occasionally. This happens at so-called *shunting yards*, which are locations with many tracks, usually near big train stations. An example of a railway hub, consisting of a station and its nearby shunting yard is shown in Figure 1.

Figure 1: Railway hub in Amersfoort.

Planning movements and tasks on these shunting yards is a major challenge. The problem of finding a feasible shunting plan containing the operations to route trains between a station and shunting yard, assign the parking tracks, and schedule the servicing tasks is commonly known as the **Train Unit Shunting Problem (TUSP)** [1]. It consists of the following components: matching arriving to departing train compositions, combining and splitting train units, parking incoming train units, routing train units to their allocated tracks, and scheduling of the service actions to be performed on train units. These are all NP-hard problems which, in combination, lead to a computationally extremely difficult problem.

Currently, most of the planning is still done manually by humans, which is a sensitive, time-consuming task that is prone to errors [2]. Moreover, the utilization of the shunting yards is constantly growing due to an increasing rolling stock, and because the capacity of the shunting yard is not increasing due to a lack of physical space. Therefore, many researchers have raised the need to develop automated planning tools to help human planners [1–4]. However, current existing models are not able to distinguish valuable patterns in the data, as opposed to human planners [2]. The aim of this paper is to examine train data for patterns that could be utilized in the automated planning process. This data contains train positions for different timestamps, which can be combined to reconstruct the routes of individual trains. Together, these routes construct a solution to the Train Unit Shunting Problem.

More specifically, this paper aims to answer the following research question: ***How to automatically detect whether a solution to the Train Unit Shunting Problem is a week or a weekend day?*** Current automated planning models, such as the local search algorithm proposed in [4] that is utilized by the Nederlandse Spoorwegen (NS), take as main input a timetable with arrival and departure time of trains and a description of the infrastructural layout of the shunting yard. However, they do not take into account any additional information about the respective day they are used to create a plan for. This means that they utilize the same heuristic for generating shunting plans for both week and weekend days. To address this gap, this paper concentrates on examining train data for valuable patterns over week and weekend days. Such patterns could be helpful in designing separate planning heuristics that are optimized specifically for week or weekend days. The main research sub-questions that this paper aims to answer are:

- *What is the distribution of the number of parked trains in a shunting yard during a week/weekend day?*
- *What types of trains are parked most often during a week/weekend day?*
- *What tracks are most used for shunting during a week/weekend day?*
- *For each type of train, what tracks are most used for shunting this type during a week/weekend day?*

Paper Structure

This paper has the following structure: Section 2 provides the background of the research. Section 3 presents the methodology followed in this paper. Section 4 shows the experimental setup and the obtained results. In Section 5, the ethical aspects of the research are discussed. Section 6 summarizes and discusses the findings of the research, as well as proposes future improvements. Finally, a conclusion is presented in Section 7.

2 Background

This section aims to present the background that is needed to understand the experiment procedure followed in this paper. First, a more in-depth formal introduction to the Train Unit Shunting Problem is presented. Next, the dataset used in this paper is described in detail. There follows a discussion about the frameworks and tools utilized within the experimental part of the research. Finally, there is a description of the methods used in the experiment.

2.1 Introduction to the Train Unit Shunting Problem

The problem considered in this paper is the Train Unit Shunting Problem with Service Scheduling (TUSPwSS) described in [4], which is an extension of the original Train

Unit Shunting Problem (TUSP) formulated in [5]. Throughout this paper, this problem will be referred to as the TUSP for simplicity.

The TUSP takes as input a timetable detailing the arrivals and departures of trains and a description of the infrastructural layout of the shunting yard, along with a set of resources as well as the service activities of each train unit that need to be completed before it leaves the service site [4]. The goal of the TUSP is to determine whether a viable shunting plan can be established. The problem consists of the following five components:

- **Matching**: Assigning arriving train units to unique positions in departing trains that match the required composition.
- **Combining and Splitting**: Trains might need to be split and combined to form the desired matching. Train units are classified by a type and a sub-type. Units of the same type can be combined: a train is a sequence of one or more train units. More information about types and sub-types at NS can be found in Appendix A.
- **Parking**: When not moving, trains should be parked in the shunting yard on a track that meets the capacity requirements of the corresponding train. Trains are permitted to relocate during their time in the shunting yard.
- **Routing**: A path over the infrastructure for each train movement. Train collisions or crossings are not allowed in a feasible shunting plan.
- **Service scheduling**: Service activities should be completed before the departure of the corresponding train. Service activities take place at service sides including a set of resources, such as cleaning resources or maintenance crews [4]. Each resource is restricted to operating on trains parked on designated tracks.

2.2 Description of the Dataset

The data considered in this paper is provided by ProRail and contains information about NS trains only. The data consists of train positions for different timestamps, collected from seven railway hubs throughout the Netherlands over a ten-month period, from May 2023 to February 2024. These train positions over time can be combined to reconstruct the routes of individual trains. Together, these routes form a solution to the TUSP. More information about the dataset can be found in Appendix B.

Since the raw data is obtained from inaccurate GPS trackers, ProRail has additionally provided a processed version of the data, where trains are projected on a path and corrected to belong to the same path. This makes the data more accurate and easier to work with.

The position of a train unit contains, among others, the track where the train unit is at the time of the sample. Therefore, ProRail maintains an additional dataset that contains all tracks part of all seven railway hubs, as well as information about each track such as area, type, length, etc. This enables filtering only the tracks that are part of a certain railway hub under consideration.

2.3 Frameworks and Tools

The data utilized in this study is owned by ProRail and stored in a private container on Microsoft Azure [1]. It is composed of multiple datasets for each month from May 2023 to February 2024 for each shunting yard, stored in parquet files. Moreover, ProRail provides a private Python repository containing methods for data manipulation and visualization. The data is accessed and processed using analysis and manipulation tools such as pandas [2] and polars [3].

2.4 Methods

The main research question considered in this paper examines the problem of detecting whether a solution to the TUSP is a week or a weekend day. This can be viewed as a classification task, where a solution is classified as one of two labels: weekday or weekend day. Therefore, here are presented several binary classification methods that are utilized in the experiment described in Section 3, along with their advantages and disadvantages.

Logistic Regression

Logistic regression is a supervised classification machine learning model. It is good for making predictions where the output variable is categorical. It is especially well suited for binary classification tasks, such as the presence or absence of disease [6]. A more in-depth overview of the logistic regression model and its usage can be found in [7].

The main advantage of the logistic regression method is that it is easy to implement and interpret, yet efficient in training. Additionally, it makes no assumptions about the distributions of classes in feature space. Moreover, it can interpret model coefficients as indicators of feature importance.

However, logistic regression has several disadvantages. It assumes linearity between the dependent and independent variables, which limits its ability to capture complex relationships. Consequently, it is unsuitable for solving nonlinear problems due to its inherently linear decision surface.

Support Vector Machine

Support Vector Machine (SVM) is a supervised algorithm that works best in classification problems, especially with small but complex datasets. In the context of binary classification, SVM tries to find a hyperplane that best separates the two classes, according to a predefined heuristic. There are two types of SVM: *linear* and *nonlinear*. Linear SVM is used when the data is perfectly linearly separable. In the case of binary classification, this means that the data points can be classified into two classes using a single straight line. Conversely, a nonlinear SVM is used when the data is not linearly separable. More detailed information on SVM can be found in [8] and [9].

The SVM method offers several advantages. It is effective in handling high-dimensional data and generally performs well with small datasets. Additionally, it can model nonlinear decision boundaries by mapping the data into a higher-dimensional space where it becomes linearly separable.

However, the main disadvantage of the method is that it can be sensitive to the choice of parameters, such as the regularization parameter. Thus, it can be difficult to determine the optimal parameter values for a given dataset.

Random Forest

Random forest is a machine learning algorithm that combines the output of multiple decision trees to produce a single outcome. It is good for handling complex datasets and mitigating overfitting and works well for both classification and regression tasks. Detailed information about the random forest algorithm can be found in [10].

The advantages of this approach include its ability to model complex, nonlinear relationships between features

[1] https://azure.microsoft.com/en-us

[2] https://pandas.pydata.org/

[3] https://pola.rs/

and target variables, achieving high accuracy. Additionally, it provides valuable insights into the importance of each feature in the data.

However, the random forest method has several disadvantages. It can suffer from overfitting when the model captures noise in the training data and can be computationally expensive, particularly with large datasets.

3 Methodology

The aim of this section is to set up the methodology used to approach the main research question. This research paper follows an experimental procedure whose aim is to answer the research sub-questions. It consists of three steps: data preprocessing, examining the data for patterns, and binary classification of solutions to the TUSP.

3.1 Data Preprocessing

As already mentioned, the data used in this paper consists of train unit positions for different timestamps. These positions could be combined to reconstruct the routes of individual train units. Together, these routes constitute a solution to the TUSP. Therefore, as a data preprocessing step, the team had to implement a data structure that represents such a solution. More information about the solution data structure can be found in Appendix B.2.

Apart from the train unit data, a dataset containing all the tracks from all seven railway hubs is provided. It enables accessing only tracks from a particular area. However, in the experiment followed throughout this paper, a means to access only tracks from a certain shunting yard is needed. Therefore, a method is implemented that, given the name of a shunting yard in a particular area, filters only the tracks that are part of this shunting yard. By obtaining the tracks of a single shunting yard, only train units that pass through this shunting yard can be filtered from the whole data. This is useful for the experiments described in the later subsections.

3.2 Pattern Extraction

Since train positions for different timestamps have been recorded for seven shunting yards over the span of ten months, the amount of data is considerably large. Therefore, due to the limited amount of time, only data from a single area and shunting yard is considered for thorough examination for patterns.

Humans are good at identifying patterns when they see them, though, it is hard to construct a precise definition of a pattern. In [2], it is suggested that the visual pattern recognition of humans is based on innate knowledge, individual learning, and context. In the context of shunting planning, this paper extends this notion by formulating the following definition of a pattern:

Definition 1. *A pattern is a recurring or discernible regularity, trend, or arrangement within a dataset.*

In the pattern extraction part, seven random dates for each day of the week are considered. This way, if a recurring trend is found in these random days, it could be concluded with a high probability that there is a strong inherent pattern in the data. Moreover, these days are considered in groups: the weekdays (Mon–Fri) are studied for patterns that are different from patterns for the weekend days (Sat–Sun). The types of patterns that data is examined for correspond to the four research sub-questions presented in Section 1.

Finding potential patterns in the data that differ between week and weekend days could be helpful in designing new heuristics for automatically creating shunting plans.

3.3 Binary Classification of TUSP Solutions

The main research question that this paper aims to answer is how to automatically detect whether a solution to the TUSP is a week or a weekend day. Therefore, apart from examining the data for patterns in data that differ between week and weekend days, the research question can be viewed as a binary classification task on TUSP solutions. Therefore, this part of the experimental procedure is concerned with considering the trade-offs between existing binary classification methods and comparing their performance on the data. The supervised binary classification methods considered for the task are presented in Section 2.4.

4 Experimental Setup and Results

This section discusses the experimental setup that follows the methodology presented in Section 3 and presents the obtained results. It is divided into two parts. First, Section 4.1 discusses the extraction of patterns from the train data. Then, Section 4.2 presents the binary classification task performed on TUSP solutions.

4.1 Pattern Extraction

Due to the large amount of recorded train data and the limited amount of time, it was decided to consider only data from a single area and shunting yard. The area chosen in this experiment is Amersfoort, with the examined shunting yard being Amersfoort Bokkeduinen.

The randomly generated days for each day of the week are as follows:

- **Monday**: 13.11.2023
- **Tuesday**: 17.10.2023
- **Wednesday**: 20.09.2023
- **Thursday**: 08.06.2023
- **Friday**: 24.11.2023
- **Saturday**: 02.12.2023
- **Sunday**: 28.01.2024

For the purposes of this study, each day is defined as the 24-hour period from 00:00:00 to 23:59:59, thereby encompassing the entire time frame of the respective date.

Next, the raw train data from the Amersfoort area is filtered by time for each of these randomly generated days. Data records are considered to be part of the filtered data if and only if the period enclosed by the TimeStart and TimeEnd (refer to Appendix B.1) of the respective data records overlaps with the time period defined by the chosen day of the week. As a result, seven filtered datasets for each day of the week are obtained, that contain raw data from the Amersfoort area.

Then, the raw data is transformed into the solution data structure, in order to extract for each train, its path through the Amersfoort area. As mentioned earlier, since the data provided by ProRail is based on GPS recordings, it is not 100% accurate. Sometimes, glitches are observed even in the processed version of the data when visualizing it, e.g, a train traveling on track X suddenly changes its location to a nearby track Y for a single time frame and then appears back on track X. Such glitches can be problematic when considering train units in a particular shunting yard at a given point in time. To ensure that such glitches are ignored when working with data from shunting yards, a train is considered to have been in a shunting yard at least once during its path through the respective area if and only if this train has traveled on a track that is part of the shunting yard being examined for at least three consecutive time frames. As a result, there follows a filtering step to obtain from all trains in the Amersfoort area during the respective day, only those that have travelled on any track X part of the Amersfoort Bokkeduinen shunting yard for at least three consecutive time frames.

Next, the filtered data for each day is examined for patterns according to the aspects described in Section 3.2.

Distribution of the Number of Parked Train Units in the Shunting Yard During the Day

The filtered data containing only trains passing through the Amersfoort Bokkeduinen shunting yard during the respective day is then used to extract a distribution of the number of parked trains in the shunting yard throughout the day.

First, the grouped path data is converted back to raw data with positions against timestamps and is filtered to contain only entries with a value of Opgesteld for the ActivityType (refer to Appendix B.1). This way, positions are obtained only for parked trains. Next, the whole day is divided into 24 time windows for each hour of the day, and for each window, the number of parked units within that window is calculated. It is worth noting that a train unit could be counted as parked over several one-hour windows in case the unit is parked for more than an hour. Moreover, a train unit could be counted as parked more than once in the same time window in case the same unit undergoes consecutive shunting (parking) operations shortly after another.

Here are presented the obtained results for the distribution of the number of parked trains in Amersfoort Bokkeduinen over Monday and Sunday. The results for all days of the week can be found in Appendix E.1.

Figure 2 indicates that on Monday, November 13, 2023, the Amersfoort Bokkeduinen shunting yard experiences its lowest occupancy during the time intervals of 07:00–09:00 and 17:00–18:00. These intervals correspond to the peak hours. Thus, the observed minimal utilization could be explained by the fact that during peak hours, a significant amount of people commute to and from work. Conversely, the shunting yard experiences higher occupancy during the time intervals of 00:00–06:00 and 23:00–24:00, with the highest peaks occurring between 01:00 and 05:00. This could be explained by the fact that at night there is little demand and, consequently, less trains scheduled for travel, which leads to most of the trains being parked in the shunting yard. Lastly, during off-peak hours, the distribution of the number of trains parked in the shunting yard is rather similar.

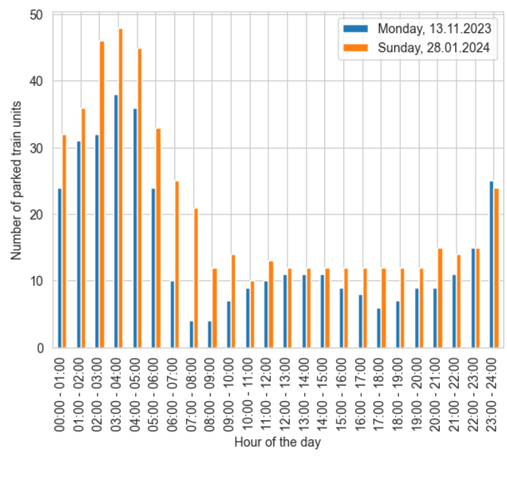

Figure 2: Distribution of the number of parked trains in the Amersfoort Bokkeduinen shunting yard

On the other hand, on Sunday, January 28, 2024, the distribution of the number of trains parked in the shunting yard is relatively similar during day hours (07:00–23:00). This could be explained by the fact that typically, in weekends, there are no clearly identifiable peak hours when there is high demand, probably due to people not commuting to and from work, but rather traveling for recreational purposes. Conversely, during night hours (00:00–07:00 and 23:00–24:00), the shunting yard experiences higher occupancy, with the highest peaks being between 02:00 and 05:00. Similarly to Monday, November 13, 2023, this could be explained by the low demand during night hours.

Number of Parked Train Units per Train Unit Type

The data containing the routes of trains passing through the Amersfoort Bokkeduinen shunting yard during the respective day is grouped by Unit Type (refer to Appendix B.1) in order to obtain the total number of parked train units for each unique unit type. Here are shown the results for Tuesday and Saturday. The results for all days of the week can be found in Appendix E.2. More information about the unit types presented in the results is available in Appendix A.

Figure 3 suggests that on Tuesday, 17.10.2023, the majority of train units parked in the Amersfoort Bokkeduinen shunting yard are of the SNG-III and SNG-IV types. This could be explained by the fact SNG are Sprinter trains (refer to Appendix A), which means that they are used for short distances, and, consequently, they may be parked and serviced more often. In contrast, the relatively small number of parked VIRM-VI and DDZ-VI trains may be attributed to the fact that VIRM and DDZ are Intercity trains designed for long-distance travel, resulting in less frequent parking and servicing.

On the other hand, on Saturday, 02.12.2023, the predominant train unit types parked in the Amersfoort Bokkeduinen shunting yard were SNG-III, SNG-IV, and SLT-VI. Similarly, this may be due to the fact that SNG and SLT are Sprinter trains utilized for short-distance travel, leading to more frequent parking and servicing. Conversely, the VIRM-IV and VIRM-VI unit types were the least represented, likely because VIRM trains are Intercity trains intended for long-distance travel, resulting in less frequent parking and servicing.

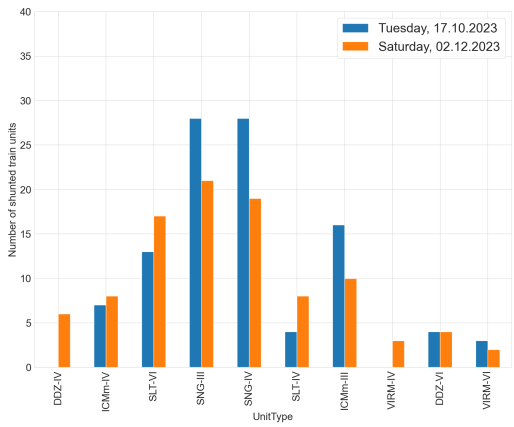

Figure 3: Number of parked trains per unit type in the Amersfoort Bokkeduinen shunting yard

Number of Parked Train Units per Track

In this step, first, the list of all tracks in the Amersfoort Bokkeduinen shunting yard is obtained. Then for each track, each train unit passing through the shunting yard is checked if it shunts (parks) on the respective track throughout its journey in the Amersfoort area. Results are shown for

Wednesday and Sunday. The results for all days of the week can be found in Appendix E.3.

Figure 4 suggests that on Wednesday, 20.09.2023, the most used track for parking within the Amersfoort Bokkeduinen shunting yard is 401R. In contrast, on Sunday, 28.01.2024, tracks 399R and 401R are most used for parking within the Amersfoort Bokkeduinen shunting yard.

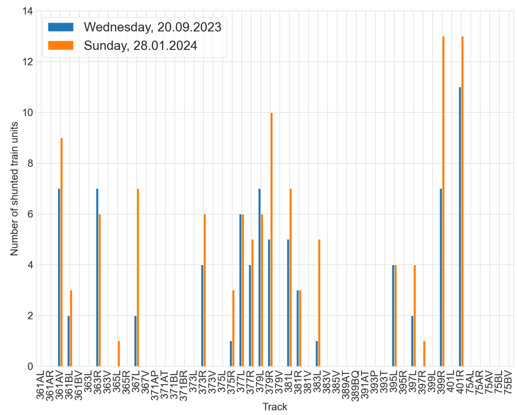

Figure 4: Number of parked trains per track in the Amersfoort Bokkeduinen shunting yard

Number of Parked Train Units over Tracks per Train Unit Type

Using the data from Section 4.1, for each unit type, the distribution of the tracks used for parking train units of the respective unit type is calculated. The data is normalized so that the percentage proportion of the number of parked units on a particular track is depicted. Moreover, sub-types are aggregated, so that the main types are only considered. Shown here are the results for Thursday (Figure 5) and Saturday (Figure 6).

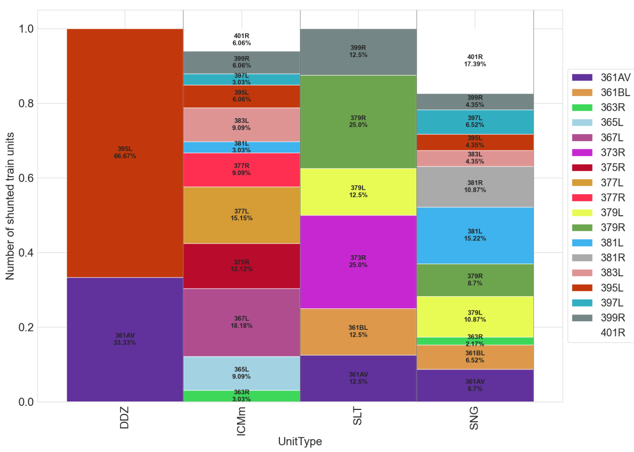

Figure 5: Distribution of the Number of Parked Train Units over Tracks per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Thursday, 08.06.2023

Figure 5 suggests that on Thursday, 08.06.2023, train units of type DDZ use only two tracks for parking, namely 395L and 361AV, which indicates a consistency in the choice of planners to park this train type on specific tracks. For type SLT, 50% of the train units are parked on tracks 379R and 373R. For type ICMm-IV, almost 43% of the train units favor track 367L, whereas the rest of the train units are similarly distributed over other tracks. The remaining train types are distributed over a bigger set of tracks and no consistency in the choice of certain tracks for parking can be observed from the data.

On the other hand, Figure 6 indicates that consistency in the choice of planners for parking tracks for type VIRM can be observed, due to the small set of tracks used for parking. Moreover, for types, DDZ, ICMm, and SLT, it can be observed that a significant part of train units are parked on a single track (361BL, 367L, and 373R, respectively), whereas the rest of the train units are relatively equally distributed over several other tracks. Finally, units of the rest of the train types are relatively evenly distributed over a larger set of tracks and, consequently, no consistency in the choice of parking tracks can be derived from the data.

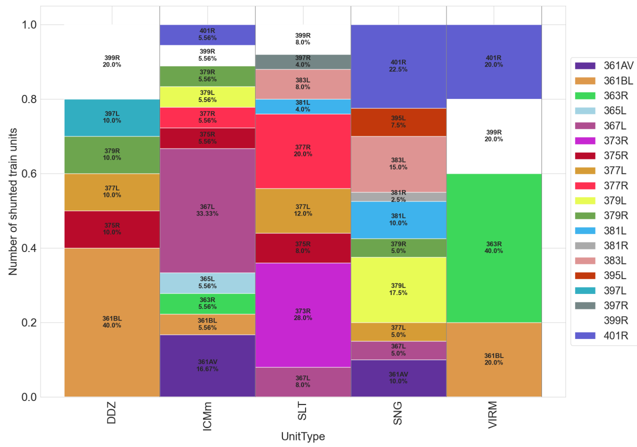

Figure 6: Distribution of the Number of Parked Train Units over Tracks per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Saturday, 02.12.2023

4.2 Binary Classification of Solutions to the TUSP

Analogous to the pattern extraction procedure, in this part of the experimental research data only from the Amersfoort area is considered for obtaining solutions to the TUSP for classification, with the Amersfoort Bokkeduinen being the examined shunting yard.

It is worth mentioning that there is a slight change in the setup in comparison to the pattern extraction part of the experiment. In a discussion with a researcher at NS, they kindly provided valuable insights about the planning procedure at NS: a solution for a weekday is obtained by considering train units and their respective schedules within the time frame between 08:00 of the same day, and 08:00 of the next day, exclusive. The reason is that shunting yards are typically the least full around 08:00 when most of the rolling stock is used to supply the peak passenger demand, which serves as a clear starting point for a given day. Conversely, a solution for a weekend day is established by considering the time window between 08:00 on a Saturday, and 08:00 on a Monday, exclusive. This is due to the fact that planners expect the shunting yards to experience consistent levels of activity over the weekends. Moreover, as opposed to the weekdays, fewer train units are used during weekends days, which suggests that more trains are parked in the shunting yard. It is also not uncommon that certain train units are only used during the week and are parked over the whole weekend until the next week. That said, planners at NS expect that less complexity is imposed on the planning process during weekends and they prefer to consider shunting plans for the weekend as a whole. Therefore, instead of using the time window between 00:00:00 and 23:59:59 for each day, solutions of the TUSP are obtained in accordance with the aforementioned procedure proposed by NS.

The procedure for the data preprocessing required for this part of the experiment is described in Appendix C. Following this procedure, 44 solutions are obtained for each of Mondays, Tuesdays, Wednesdays, and Thursdays, 43 solutions for Fridays, and 43 solutions for weekends, resulting in a total of 262 solutions for the Amersfoort area over the period from May 1, 2023, to February 29, 2024. One of them is dropped due to the presence of NaN values for some parameters. The rest of the solutions are labeled as belonging to a week or a weekend day by adding a *Day Type* parameter, that takes the values 1 (weekday) and 0 (weekend day).

Before employing the binary classification methods on the TUSP solutions, the data is first analyzed. Figure 7 shows the distribution of the target variable, Day Type, within the solution dataset. It can be observed that the entries of class 1 (weekday) are more than the entries of class 0 (weekend day), which indicates that there is a class imbalance.

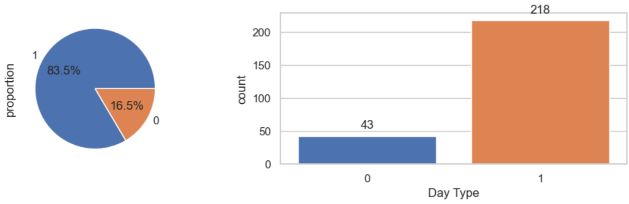

Figure 7: Target variable distribution of the solution dataset.

Next, feature scaling is performed on the solution data in order to ensure numerical stability, and prevent features with larger scales from dominating the model, thereby improving the overall model performance and interpretability. Standardization is utilized as the scaling technique, so that features have zero mean and unit standard deviation. The dataset is split into a training and test dataset, containing 70% and 30% of the total data, respectively.

Next, the solutions are used as input to the binary classification methods presented in Section 2.4. Here are presented the results obtained by the logistic regression method. The results for the other methods can be found in Appendix D.

Logistic Regression

A logistic regression model is created, where the class imbalance problem in the data is addressed by adding class weights to the model. The purpose of incorporating such class weights is to penalize misclassifications of the minority class by assigning it a higher weight, while correspondingly reducing the weight for the majority class. The model's performance is further improved by performing a hyperparameter optimization using the *grid search algorithm*. The algorithm works as follows: all possible values for each hyperparameter are specified in a predefined grid. Then, the model is trained and evaluated for each possible combination of hyperparameters in the grid, applying the cross-validation approach to ensure reliable results. The performance of each model is subsequently compared, and the combination of hyperparameters yielding the optimal performance is selected.

Figure 8 shows the results of employing the logistic regression model on the test dataset using the optimal hyperparameters. Here, *F1-score* is utilized as a performance metric in order to address the class imbalance problem in the dataset. It provides a harmonic mean of *precision* and *recall*, thereby balancing the trade-off between these two metrics [11]. Precision measures the accuracy of the positive predictions, while recall assesses the ability to identify all positive instances. In imbalanced datasets, the F1-score ensures that both false positives and false negatives are considered, offering a more comprehensive evaluation of the model's performance compared to accuracy, which can be misleading due to the predominance of the majority class.

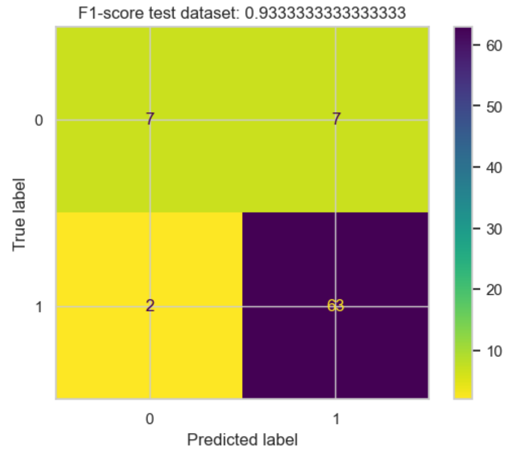

Figure 8: Confusion matrix and F1-score of the logistic regression model.

Figure 8 suggests that the F1-score of the model is relatively high, which indicates that the model has a well-balanced performance, combining high precision and high recall. Moreover, the confusion matrix shows that the majority of the test samples with label 1 are indeed classified as class 1. Nevertheless, only half of the test samples with label 0 are classified as class 0.

5 Responsible Research

This section describes the steps taken in order to guarantee that the research presented in this paper is done responsibly. First, the transparency and reproducibility of the research is described. Next, the research integrity is discussed.

5.1 Research Reproducibility

It is essential for research to be reproducible so that future scientists can effectively build upon the ideas of earlier generations. In order to ensure this, the code used within the experimental part of the research is publicly available. The files, variables, and methods have meaningful names to further help interested readers understand the code more easily. Nevertheless, the code operates on proprietary data held by ProRail, which means that the experiment described in this paper could not be fully replicated without this data. However, upon request, ProRail may provide interested researchers access to the data.

Furthermore, no special technical specifications are required in order to reproduce the experiment described in this paper.

5.2 Research Integrity

This research has investigated the presence of interesting properties in train unit data provided by ProRail. The results obtained from the experiment could be used in the sake of improving the automated planning process in the context of the TUSP and making it more efficient. Nevertheless, readers should be aware of the time limitations that deem the results to be not directly applicable in practice without further research.

6 Discussion

This section is going to elaborate on the results presented in Section 4. Section 5.1 discusses the pattern extraction part. The results from the binary classification part are examined in Section 5.2.

6.1 Pattern Extraction

Here the results obtained for each of the research subquestions are presented in more detail.

Distribution of Shunted Train Units in the Shunting Yard During the Day

In a broader context, Figures 11–15 suggest that a similar distribution of the number of shunted trains in the Amersfoort Bokkeduinen shunting yard can be observed for the weekdays Mon–Fri. This distribution is characterized by low occupancy during peak hours, high occupancy during night hours, and moderate, consistent occupancy during off-peak hours. This behavior can be attributed to the fact that weekdays are mostly working days, leading to increased commuter activity during peak hours and minimal passenger travel during night hours. Consequently, this pattern reflects the typical distribution of parked trains in the shunting yard throughout the weekdays.

Similarly, Figures 16 and 17 show a comparable distribution in the number of shunted trains over the weekend, with high utilization during night hours and lower, consistent utilization during the day. This may be attributed to increased leisure travel during the day on weekends. Thus, this suggests a pattern in the number of parked trains on weekends.

To gain further insights and identify meaningful patterns, new research directions should be pursued. For instance, an aspect that could be taken into account is whether a day is a public holiday or a non-working day. The randomly drawn dates used in the experiment are all non-holiday days by chance. It is important to note that on non-working days and public holidays, the distribution of parked trains throughout the day will almost certainly differ from that on working days, likely resembling the distributions observed on weekend days.

Additionally, weather forecasts could be utilized to identify patterns in the distribution of parked trains. For example, one could expect that on a rainy working day, more people travel to and from work by train instead of cycling, which would result in occupancy in the shunting yard during peak hours that is lower than usual. Moreover, on a weekend day, if the weather is bad, people are expected to travel less. Conversely, if the weather is good, more people are expected to enjoy traveling. Such hypotheses could be tested against data enriched with details about the weather forecast.

Number of Shunted Train Units per Train Unit Type

Figures 18–24 collectively show that SNG train units are consistently the most parked type in the Amersfoort Bokkeduinen shunting yard across all days of the week. This may be attributed to the fact that SNG are Sprinter trains used for short-distance travel, necessitating more frequent parking and servicing. Furthermore, as indicated in Table 1 of Appendix A, there are 205 train units of type SNG, establishing it as the most prevalent type and likely the most frequently utilized. However, no patterns for the number of parked train units of type SNG can be deduced that differ between week and weekend days.

Moreover, the figures indicate that train units of type DDZ are parked less frequently in the Amersfoort Bokkeduinen shunting yard across all days of the week. This could be explained by the fact that DDZ are Intercity trains used for long-distance travel and, therefore less frequently parked and serviced. Moreover, according to Table 1, there are 49 train units of type DDZ, which makes it the least prevalent type and probably less frequently utilized. It should be noted that Figure 24 indicates the absence of DDZ-type train units parked in the shunting yard on Sunday, January 28, 2024.

For the other train types, no general observations could be yielded, due to the absence of regularity in the number of parked train units of the respective type over the different days of the week. As a consequence, no discernible patterns in the number of shunted trains per unit type between weekdays and weekends can be identified.

In order to extract more meaningful insights from the data, additional aspects must be taken into account, such as the current state of the rolling stock. For practical and financial reasons, trains are not purchased all at once but rather over different time intervals [2]. Therefore, at any point in time, the number of train units per type may differ from the figures presented in Table 1. Utilizing this information could assist in identifying the reasons for the low or high numbers of parked train units of a particular type during an observed day.

Furthermore, rather than analyzing the number of parked train units per unit type daily, studying the data on an hourly basis could be advantageous. This approach could facilitate the identification of low-level patterns in the data, such as specific train types being more frequently utilized during certain times of the day, such as in the afternoon. Thus, a pattern of the utilization of the shunting yards could be learned from the data.

Number of Shunted Train Units per Track

In a wider perspective, Figures 25–31 indicate a consistent usage of specific tracks for parking train units at the Amersfoort Bokkeduinen shunting yard throughout the week. Some of the most utilized tracks within this set are 401R, 361AV, 379L, 379R, 381R, and 363R. This could be explained by the specific infrastructural layout of the Amersfoort Bokkeduinen shunting yard [4]. The layout suggests that there are only two entry/exit tracks, making them critical bottlenecks through which every train unit must pass upon entering or leaving. According to a researcher at NS, who provided valuable insights for this research paper, this is the reason why planners avoid parking train units on these tracks, as this could lead to collisions of trains and sub-optimal shunting plans. Consequently, the set of tracks used for parking mentioned earlier comprises tracks exclusively dedicated to parking and servicing, not for entering/exiting the shunting yard. Therefore, this accounts for the similarity in parking track selections across different days of the week. As a result, no discernible patterns could be derived from the observed data that are specific for week and weekend days.

Number of Shunted Train Units over Tracks per Train Unit Type

In a broader perspective, Figures 32–38 suggest that during the randomly drawn weekdays, train units of type DDZ are consistently parked on a small set consisting of 2–4 tracks within the Amersfoort Bokkeduinen shunting yard. However, on Saturday, December 2, 2023, DDZ units occupy a larger set of six tracks, some of which are not typically used for weekday parking. Moreover, on Sunday, January 28, 2024, no DDZ units are parked in the yard. This observation can serve as a pattern that indicates planners prefer a specific set of tracks for parking DDZ units on weekdays, whereas on weekends, there is less adherence to a fixed pattern, possibly due to the availability of a higher number of

[4]https://sporenplan.nl/ provides detailed information about the track plans and layouts of shunting yards in the Netherlands.

less busy tracks. This preference aligns with the nature of DDZ trains used for long-distance travel, which are parked less frequently and likely benefit from longer-term parking on quieter tracks during weekdays. Conversely, during weekends, when generally fewer trains are utilized, planners may distribute the DDZ train units more freely within the less busy shunting yard.

Additionally, VIRM train units are typically parked on a subset of 2–5 tracks, both on weekdays and weekends. Similar to DDZ trains, this is likely because VIRM trains, used for Intercity travel, are parked less frequently and require longer-term parking on specific, less busy tracks within the shunting yard.

In contrast, train units of other types are typically parked on a larger variety of tracks, both on weekdays and weekends. For Sprinter trains like SLT and SNG, this could be due to their frequent use for short-distance travel, resulting in more regular parking needs. Therefore, planners likely assign these train units to any available parking track that can accommodate them without disrupting other operations, while ensuring a feasible shunting plan. Overall, there are no clear patterns observed for train types other than DDZ that distinguish between weekdays and weekends.

Conclusions

The methodology employed for pattern extraction focused on seven randomly selected days from a ten-month dataset due to its extensive volume. While this approach reveals the presence or absence of examined patterns, it does not definitively establish their presence or absence in the entire dataset. Therefore, general conclusions about the existence or nonexistence of patterns in the whole data require further analysis.

The proposed recommendations aim to enhance the automated planning process for shunting operations. Currently, automated planning models primarily rely on a timetable detailing train arrival and departure times, along with the layout of the shunting yard infrastructure. However, incorporating additional factors such as weather forecasts and holiday schedules could enable the models to employ more tailored heuristics for week and weekend days. This approach would optimize shunting plans to better align with passenger demand and improve the efficiency of rolling stock utilization. Moreover, to effectively predict and schedule servicing tasks for different train types, a deeper analysis of shunting yard utilization patterns based on comprehensive data across various days and times is essential. Therefore, a thorough examination of the entire dataset is crucial to identify and validate any recurring patterns in the distribution of shunted train units across tracks.

6.2 Binary Classification of Solutions to the TUSP

The results of the binary classification of solutions to the TUSP, presented in Figures 8–10, demonstrate that the models effectively classify weekday solutions, while achieving high F1-scores that indicate balanced performance in terms of precision and recall. However, they do not consistently achieve accurate classification of weekend solutions.

The logistic regression model's imperfect performance may stem from its assumption of linearity, which is inadequate for capturing the likely non-linear relationships in the high-dimensional data used in this experiment. The support vector machine model may underperform due to class imbalance and a small number of training samples, as well as potential overlap in target classes. Lastly, the random forest model's poor performance could be attributed to noise in the data from inaccurate GPS trackers, reducing its generalization ability.

The original data representing train paths through the Amersfoort area is extensively modified to serve as input for the binary classification models used in this experiment. This modification includes averaging each train unit's path and aggregating all train units into a single record, which diminishes variability and obscures finer details and potentially meaningful patterns. Additionally, the data is only augmented with one extra feature: the number of parked trains in the Amersfoort Bokkeduinen shunting yard at each timestamp. To gain further insights, it would be beneficial to extract and utilize more complex features in the binary classification problem.

In summary, Figures 8–10 show that the models used for binary classification of TUSP solutions achieve balanced performance and accurately classify many test samples, despite limited data optimization and preprocessing. This suggests that the data contains intrinsic patterns distinguishing weekday from weekend solutions. Enhanced performance could be achieved with more thorough preprocessing, data optimization, and advanced classification methods. These improvements could inform real automated planning by incorporating data-driven patterns and developing distinct heuristics for weekdays and weekends, optimizing shunting plans.

6.3 Finishing Words

The results presented from the pattern extraction and binary classification experiments are based on a limited dataset due to time constraints, focusing specifically on the Amersfoort area and the Amersfoort Bokkeduinen shunting yard. It is crucial to recognize that outcomes could vary significantly for different areas and shunting yards due to factors such as infrastructural layout, area traffic, and geographic distinctions. Therefore, separate analyses for various locations and shunting yards are essential.

As detailed in Section 5, the experimental methodology employed in this study ensures reproducibility. Hence, similar experiments can be conducted for any area and shunting yard of interest. The resulting insights can be leveraged to enhance the automated shunting process in respective areas by refining model heuristics tailored specifically to each location.

7 Conclusions and Future Work

In this paper, realization data containing train unit positions for different timestamps is examined for the existence of valuable patterns that could be applied in the optimization of the automated planning process in the context of the Train Unit Shunting Problem (TUSP). First, a detailed introduction to the TUSP is given, as well as a description of the utilized data. Then, the data is analyzed for the presence of specific patterns that could be helpful to differentiate whether a solution to the TUSP obtained from the data belongs to a week or a weekend day. It is discovered that, with high probability, there exists a pattern in the distribution of the number of parked trains in a specific shunting yard that differs between week and weekend days. This pattern serves as a potential property that could help differentiate TUSP solutions and possibly design specific planning heuristics that are tailored for generating plans for different types of days. Moreover, solutions to the TUSP are extracted from the data, and several binary classification methods are utilized to categorize them as either weekday or weekend day solutions. The results indicate that the employed methods achieve good performance, which suggests the presence of underlying patterns in the data that facilitate the distinction between weekday and weekend solutions to the TUSP.

Future work might need to further examine the data in more detail in order to prove or disprove the findings presented in this paper. While in this experiment only the provided data has been utilized, subsequent research might consider using supplementary data, such as the weather forecast for a given day or whether it is a public holiday or a non-working day. This could provide further insights into the data and help in designing new heuristics for the automated planning process.

References

[1] R. Freling, R. Lentink, L. Kroon, and D. Huisman, "Shunting of passenger train units in a railway station," *Transportation Science*, vol. 39, pp. 261–272, May 2005.

[2] L. van de Gevel, "How human knowledge can support algorithmic decision-making in the train unit shunting problem - an exemplary study." 2022.

[3] I. K. Hanou, S. Dumancic, M. de Weerdt, P. van der Voort, R. van den Broek, and M. van den Akker, "Increasing the capacity of shunting yards within the current infrastructure: A computational perspective,"

[4] R. Broek, H. Hoogeveen, J. Akker, and B. Huisman, "A local search algorithm for train unit shunting with service scheduling," *Transportation Science*, vol. 56, 11 2021.

[5] L. Kroon, R. Lentink, and A. Schrijver, "Shunting of Passenger Train Units: an Integrated Approach," ERIM Report Series Research in Management ERS-2006-068-LIS, Erasmus Research Institute of Management (ERIM), ERIM is the joint research institute of the Rotterdam School of Management, Erasmus University and the Erasmus School of Economics (ESE) at Erasmus University Rotterdam, Dec. 2006.

[6] T. G. Nick and K. M. Campbell, *Logistic Regression*, pp. 273–301. Totowa, NJ: Humana Press, 2007.

[7] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[8] T. Fletcher, "Support vector machines explained," *Tutorial paper*, vol. 1118, pp. 1–19, 2009.

[9] D. Boswell, "Introduction to support vector machines," *Departement of Computer Science and Engineering University of California San Diego*, vol. 11, 2002.

[10] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[11] D. Berrar *et al.*, "Performance measures for binary classification.," 2019.

A Trains at NS

The rolling stock at NS consists of many different types of trains designed for various types of travel [2]. These types can be divided into *Intercity* and *Sprinter* trains. Sprinter trains are used for shorter distances and stop at smaller stations. Intercity trains are long-distance trains that typically connect larger cities and do not stop at smaller stations. The types of trains considered in this paper, as well as their respective number of trainsets are presented in Table 1. The information is obtained from the NS Annual Report 2023.

	Train Type	**Number of Trainsets**
Sprinter trains:		
	SLT	131
	SNG	205
Total Sprinter		**336**
Intercity trains:		
	VIRM	175
	ICMm	115
	DDZ	49
Total Intercity		**339**
Total		**675**

Table 1: Number of trainsets per unit type at NS as of 31 December 2023.

Train types are further classified into sub-types. Units of the same type can be combined: a train is a sequence of one or more train units. The sub-type indicates the number of carriages in the train [4]. For example, the SLT-IV and SLT-VI consist of four and six carriages, respectively.

B Details of the Dataset

B.1 Parameters in the Data

A single entry in the data consists of numerous parameters. The most important ones are presented here:

- **Area**: The area where the position is sampled. It can be one of Amersfoort, Arnhem Goederen, Arnhem West, Carthesiusweg, Dordrecht, Hoofddorp, and Watergraafsmeer.
- **Unit Number**: The unique number of the respective train unit.
- **GroupIdHash**: This parameter uniquely represents the time when a specific train unit enters the respective area where a position sample is taken. The GroupIdHash of a train unit remains the same over the whole route of the train unit through the area.
- **Moving**: Whether the train unit is moving.
- **ActivityType**: The type of activity that the train unit is in at the respective timestamp. The most important types are *entry* (when a train unit enters an area), *exit* (when a train unit exits an area), *Short Stop* (train unit not moving for less than 30 min), and *Opgesteld* (train unit not moving for more than 30 min). Throughout this paper, a train unit having an ActivityType value of Opgesteld will be called *parked* or *shunted*. ActivityType is also related to the Moving parameter.
- **TimeStart**: The date and time when the train unit has started a respective activity.
- **TimeEnd**: The date and time when the train unit has finished a respective activity.

- **Unit Type**: The specific type of a train unit. At NS, this could be one of DDZ-IV, DDZ-VI, E1700, ICMm-III, ICMm-IV, SLT-IV, SLT-VI, SNG-III, SNG-IV, VIRM-IV, and VIRM-VI. Due to imperfections in the data, certain train units have an unidentified unit type, denoted as "Onbekend" (Dutch for "unknown"). More information about the train fleet at NS can be found in Appendix A.

B.2 Solution Data Structure

The team had to perform preprocessing of the raw data in order to implement a data structure representing a solution to the TUSP. First, the data was grouped by GroupIdHash, in order to group all the positions of a train unit within the respective area, throughout its whole path over that area. As a result, the obtained data structure contains for each train unit GroupIdHash, the unit's Unit Number, Unit Type, and a list of positions for different timestamps in the respective area, or, in other words, the route of the train unit. Together, the routes of all trains form a solution of the TUSP.

C Data Preprocessing for the Binary Classification Experiment

As mentioned in Section 2.3, the raw data provided by ProRail consists of multiple datasets for each month from May 2023 to February 2024. Consequently, the initial step involves merging these datasets to obtain the data for the entire ten-month period. Next, the data is enhanced with an additional parameter: for each entry, the number of parked trains in a specific shunting yard (here Amersfoort Bokkeduinen is considered) at the time corresponding to the entry's timestamp is recorded.

Some of the parameters in the raw data have non-numerical values, which imposes a problem for the classification task. Such parameters are ActivityType, UnitType, Moving, etc. (refer to Appendix B for more information about these parameters). Therefore, a preprocessing step is done where a numerical encoding for these parameters is performed, and non-numerical values are converted to numerical according to this encoding.

Then, the data is filtered by time according to the procedure described in Section 4.2, in order to obtain frames of raw data needed for the construction of solutions for each weekday and weekend. Each frame of data is transformed into the solution data structure so that for each train unit, its path is extracted within the time window defined by the data frame. Together, the paths of all train units form a solution of the TUSP for the respective time window, be it a weekday or a whole weekend.

A train unit path consists of multiple positions for different timestamps, as well as additional data such as the track, type of activity, direction, etc. Thus, such a path is represented as a collection of entries. However, for the purposes of the research, the representation of such paths needs to be simplified in order to be suitable for the classification task. Thus, instead of taking the whole path of a train unit, only the averaged values of all path parameters are considered.

At this point, for each weekday or weekend, a solution is obtained that contains, among other parameters, its averaged path. Nevertheless, a more compact representation of a solution is needed, so that it can be fed into a classification model as a standalone object. Therefore, the average value of all entries of a solution collection is taken to represent the solution itself.

D Additional Results of Binary Classification on Solutions to the TUSP

Support Vector Machine (SVM)

An SVM model is created and grid search is employed, similarly to the logistic regression model, in order to find the best set of hyperparameters for optimal performance on the test data. Again, the F1-score performance metric is utilized in order to address the class imbalance problem in the dataset.

Figure 9 demonstrates that the SVM model attains a high F1-score, indicative of well-balanced performance with both high precision and high recall. The confusion matrix reveals that the majority of the test samples from class 1 are correctly classified. However, a significant portion of the samples from class 0 are misclassified as class 1.

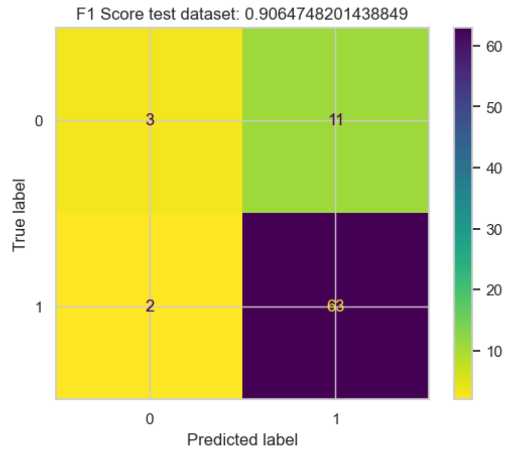

Figure 9: Confusion matrix and F1-score of the support vector machine model.

Random Forest

A random forest model is fit on the training data and hyperparameter tuning is performed using the grid search algorithm. The class imbalance problem is addressed by utilizing the F1-score as a performance metric for the model.

Figure 10 shows that the model reaches a high F1-score that indicates well-balanced performance with high precision and high recall. Furthermore, the confusion matrix suggests that all of the test samples of class 1 are correctly classified. Nevertheless, the majority of the class 0 samples are misclassified as class 1.

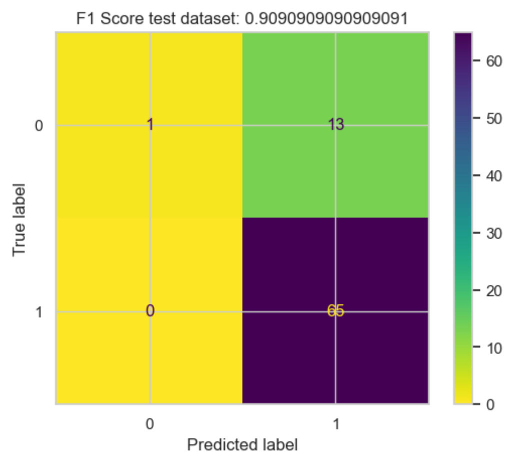

Figure 10: Confusion matrix and F1-score of the random forest model.

E Additional Figures of the Pattern Extraction Part

E.1 Distribution of parked Train Units in the Shunting Yard During the Day

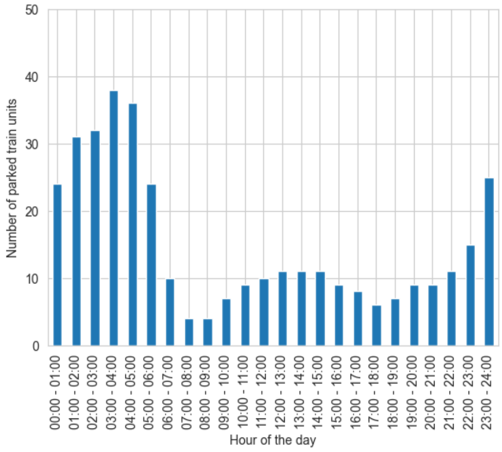

Figure 11: Distribution of the number of parked trains in the Amersfoort Bokkeduinen shunting yard - Monday, 13.11.2023

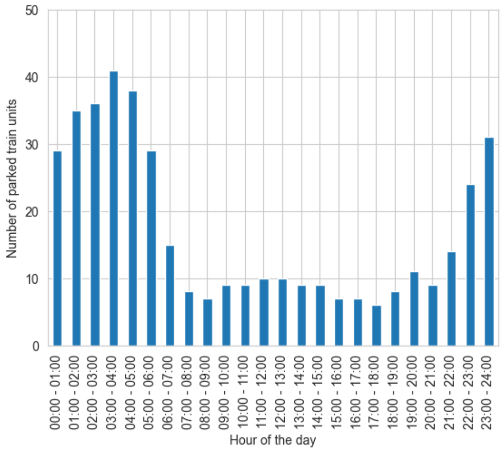

Figure 12: Distribution of the number of parked trains in the Amersfoort Bokkeduinen shunting yard - Tuesday, 17.10.2023

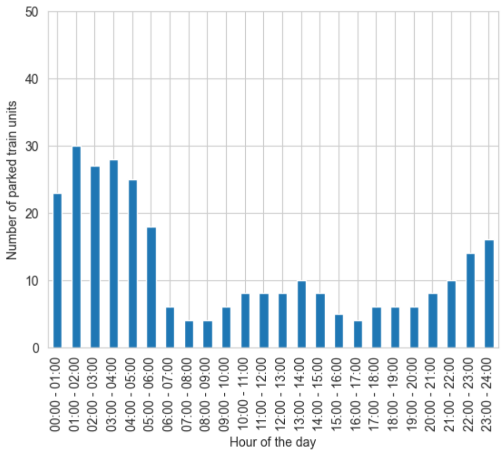

Figure 13: Distribution of the number of parked trains in the Amersfoort Bokkeduinen shunting yard - Wednesday, 20.09.2023

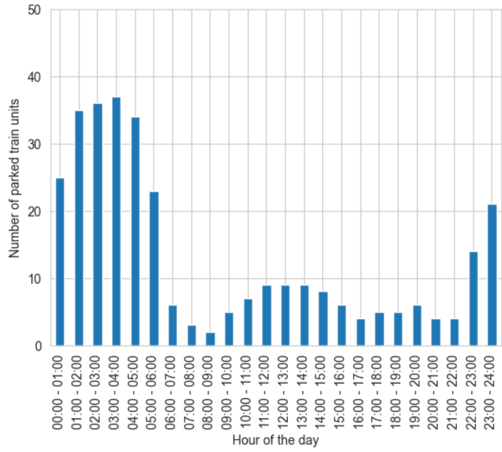

Figure 14: Distribution of the number of parked trains in the Amersfoort Bokkeduinen shunting yard - Thursday, 08.06.2023

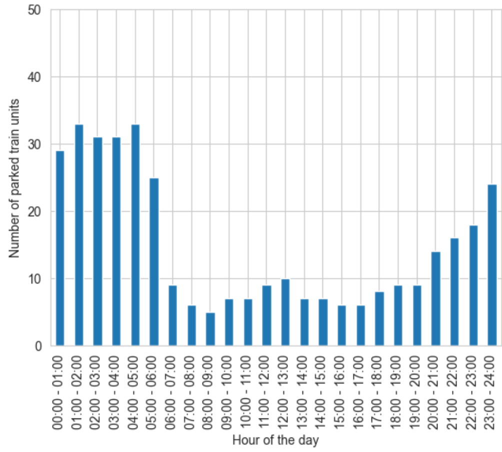

Figure 15: Distribution of the number of parked trains in the Amersfoort Bokkeduinen shunting yard - Friday, 24.11.2023

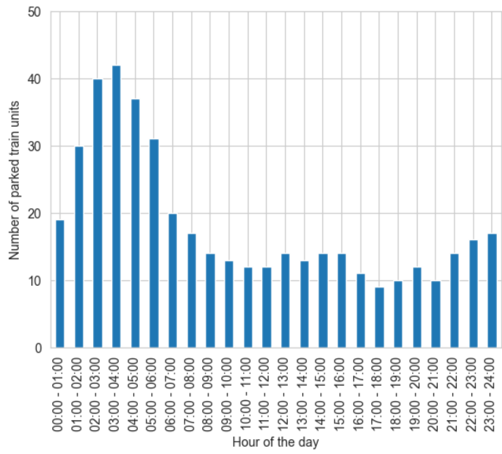

Figure 16: Distribution of the number of parked trains in the Amersfoort Bokkeduinen shunting yard - Saturday, 02.12.2023

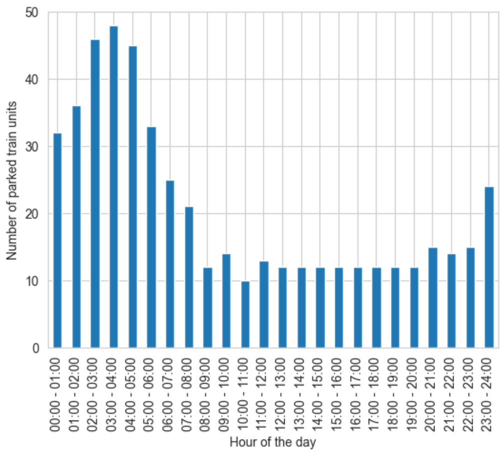

Figure 17: Distribution of the number of parked trains in the Amersfoort Bokkeduinen shunting yard - Sunday, 28.01.2024

E.2 Number of parked Train Units per Train Unit Type

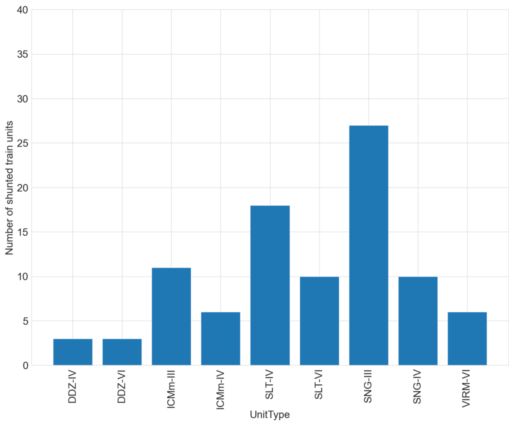

Figure 18: Number of parked Train Units per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Monday, 13.11.2023

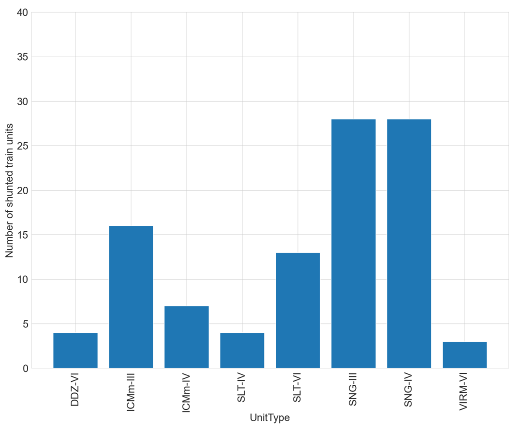

Figure 19: Number of parked Train Units per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Tuesday, 17.10.2023

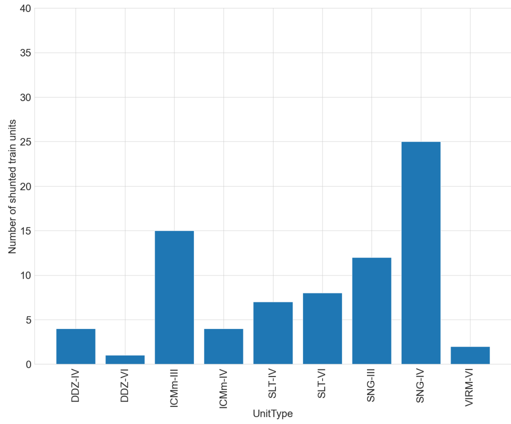

Figure 20: Number of parked Train Units per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Wednesday, 20.09.2023

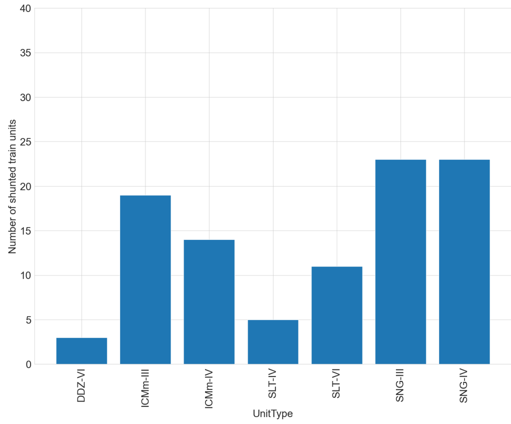

Figure 21: Number of parked Train Units per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Thursday, 08.06.2023

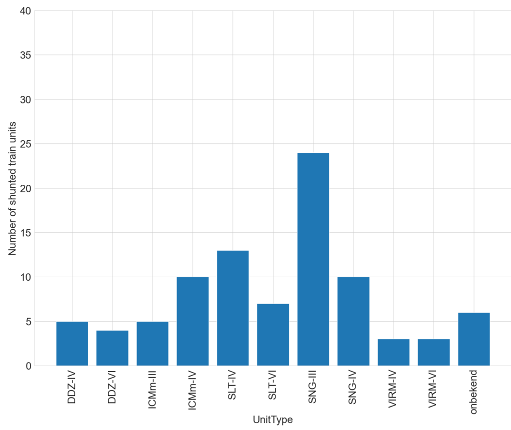

Figure 22: Number of parked Train Units per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Friday, 24.11.2023

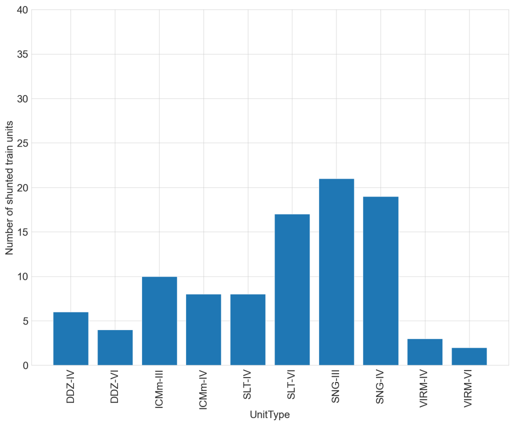

Figure 23: Number of parked Train Units per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Saturday, 02.12.2023

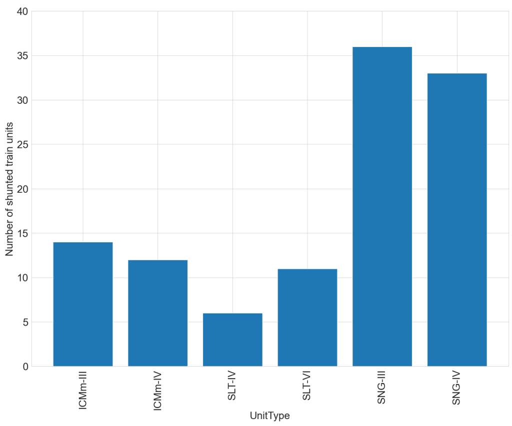

Figure 24: Number of parked Train Units per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Sunday, 28.01.2024

E.3 Number of parked Train Units per Track

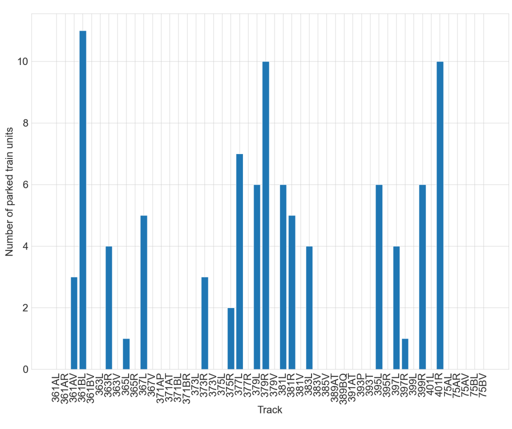

Figure 25: Number of parked Train Units per Track in the Amersfoort Bokkeduinen shunting yard - Monday, 13.11.2023

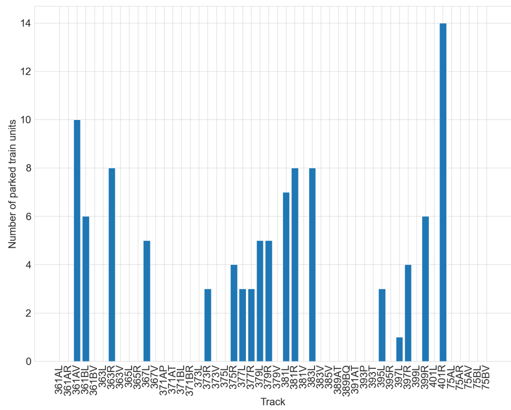

Figure 26: Number of parked Train Units per Track in the Amersfoort Bokkeduinen shunting yard - Tuesday, 17.10.2023

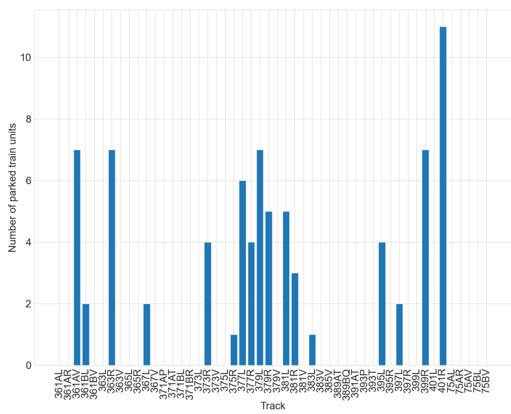

Figure 27: Number of parked Train Units per Track in the Amersfoort Bokkeduinen shunting yard - Wednesday, 20.09.2023

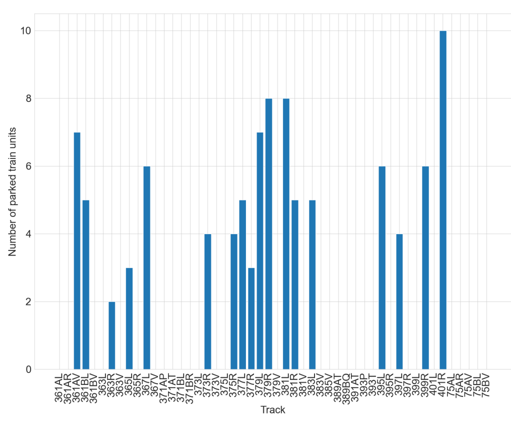

Figure 28: Number of parked Train Units per Track in the Amersfoort Bokkeduinen shunting yard - Thursday, 08.06.2023

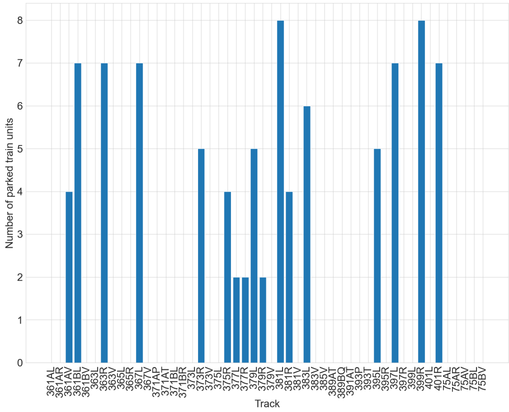

Figure 29: Number of parked Train Units per Track in the Amersfoort Bokkeduinen shunting yard - Friday, 24.11.2023

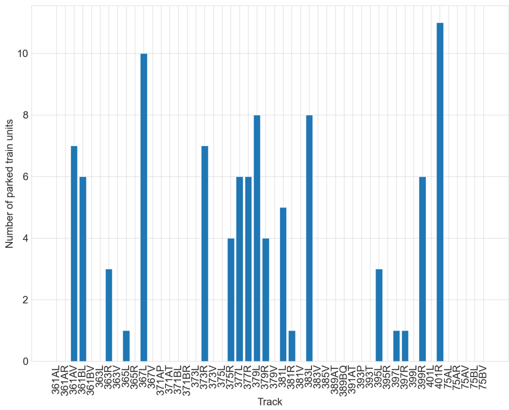

Figure 30: Number of parked Train Units per Track in the Amersfoort Bokkeduinen shunting yard - Saturday, 02.12.2023

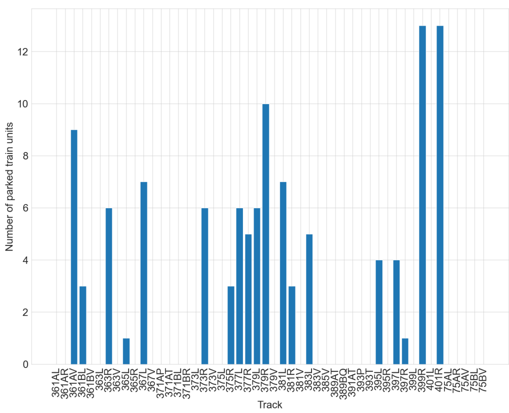

Figure 31: Number of parked Train Units per Track in the Amersfoort Bokkeduinen shunting yard - Sunday, 28.01.2024

E.4 Distribution of the Number of Parked Train Units over Tracks per Train Unit Type

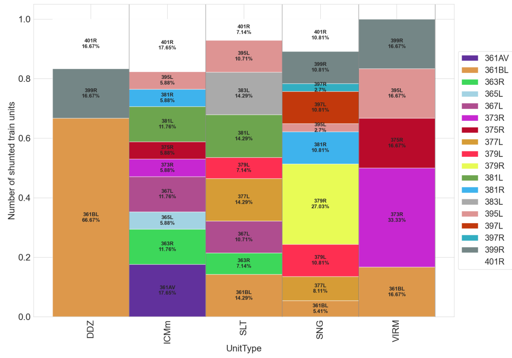

Figure 32: Distribution of the Number of parked Train Units over Tracks per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Monday, 13.11.2023

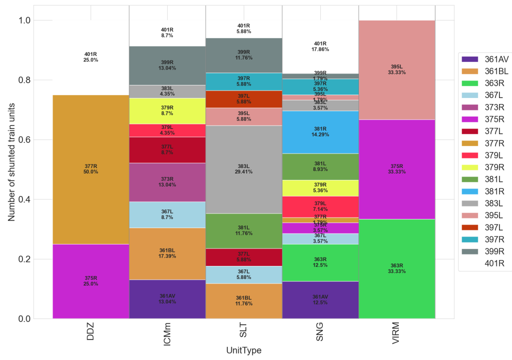

Figure 33: Distribution of the Number of parked Train Units over Tracks per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Tuesday, 17.10.2023

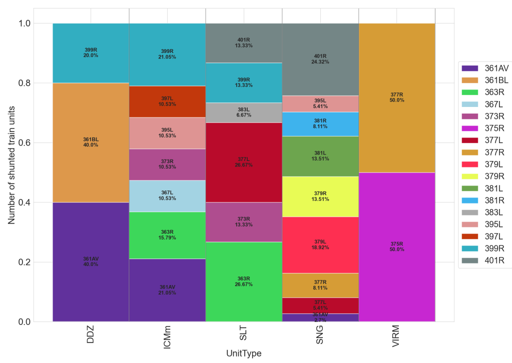

Figure 34: Distribution of the Number of parked Train Units over Tracks per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Wednesday, 20.09.2023

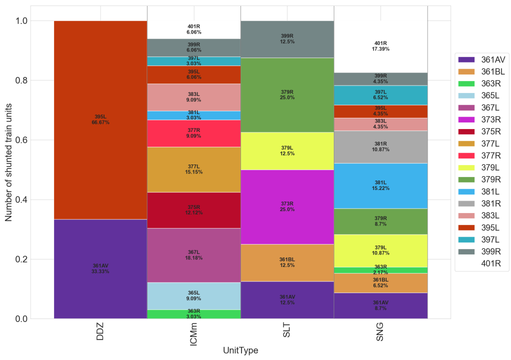

Figure 35: Distribution of the Number of parked Train Units over Tracks per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Thursday, 08.06.2023

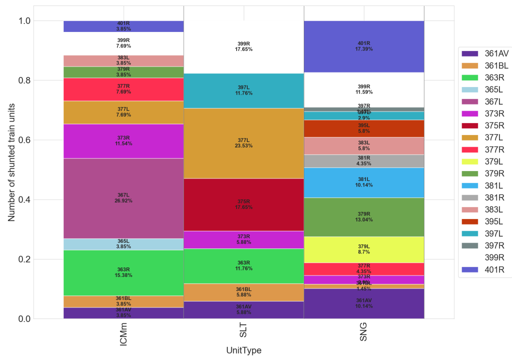

Figure 38: Distribution of the Number of parked Train Units over Tracks per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Sunday, 28.01.2024

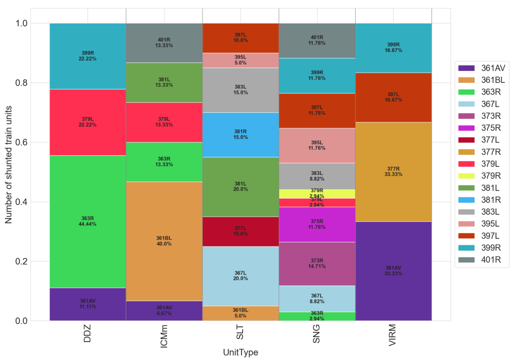

Figure 36: Distribution of the Number of parked Train Units over Tracks per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Friday, 24.11.2023

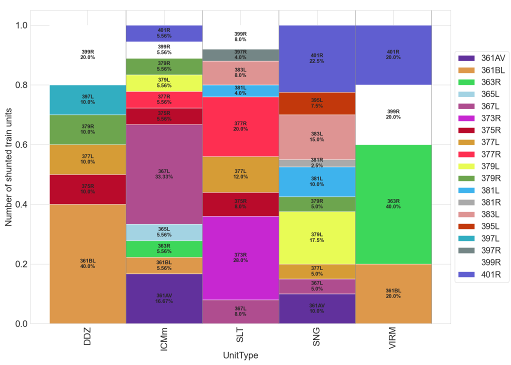

Figure 37: Distribution of the Number of parked Train Units over Tracks per Train Unit Type in the Amersfoort Bokkeduinen shunting yard - Saturday, 02.12.2023