A data-driven model for short-term prediction of arrival delay times in freight rail operations

Pineda-Jaramillo[a*], J. Bigi, F[b]. Viti, F[c]

[a] Department of Engineering, University of Luxembourg, Esch-sur-Alzette, Luxembourg
juan.pineda@uni.lu
[b] Department of Engineering, University of Luxembourg, Esch-sur-Alzette, Luxembourg
federico.bigi@uni.lu
[c] Department of Engineering, University of Luxembourg, Esch-sur-Alzette, Luxembourg
francesco.viti@uni.lu
* Corresponding author

Extended abstract submitted for presentation at the 11th Triennial Symposium on Transportation Analysis conference (TRISTAN XI)
June 19-25, 2022, Mauritius Island

March 29, 2022

Keywords: data-driven models, rail operation delays, gradient boosting, freight transport.

1 INTRODUCTION

Rail is increasingly selected for freight transport due to its advantages with respect to operational costs, efficiency, reliability, emissions, safety, and it is gradually used within the intermodal context, motivating national and international agencies to promote a shift of passengers and freight from other alternatives (Cacchiani *et al.*, 2010; Pineda-Jaramillo *et al.*, 2021). Considering these trends, rail intermodal operations play a key role in terms of reliability for the freight transport sector, and it is therefore essential to optimize all aspects including the operational use of the rail infrastructure.

Given the complexity of the rail networks and the amount of rolling stock running on it, an important aspect to consider are the delays, which can be classified in those that are caused directly by the variability of process times preparing the train for departure, and those caused by the variability in the actual operation of the train along its journey (Goverde *et al.* 2016). Arrival delay prediction (i.e., numerical difference between the scheduled arrival time and the actual arrival time in a trip between a pair of stations) is necessary because once disturbances occur, train dispatchers must assess their impact on the overall schedule and try to reduce losses by adjusting the operation, in order to diminish the chain of delays that could affect the entire system operation (Bešinović *et al.* 2016).

Although train disturbance prediction models have been developed using various approaches (De Martinis and Corman 2018; Zheng and McDonald 2021; Barbour et al. 2018), they still have difficulty in predicting the arrival delay time in the short-term once the train leaves the previous control station, and they also fail to identify the underlying causes of delay and the expected impact on operations, which significantly limits the efficacy of mitigation actions. To address these issues, the purposes and major contributions of this study are to examine different data-driven models for a short-term prediction of arrival delay time in freight rail operations, and then to examine the importance of the features associated with arrival delay time, with the aim of developing a tool to assess operational interventions to reduce disturbances and their subsequent delays in freight operations.

2 METHODOLOGY

2.1 Data collection, data pre-processing and feature engineering

Data related to freight rail operations performed between November 2019 to April 2021 were provided by the National Rail Company of Luxembourg - CFL Multimodal. Many attributes related to characteristics related to train, wagon, station and operations can be found among the collected datasets. The datasets were merged and processed to count all freight rail operations made between Bettembourg (Luxembourg) and other nine stations within the EU (Boulou, Champigneulles and Lyon in France; Zeebrugge and Antwerp in Belgium; Kiel and Rostock in Germany; Poznan in Poland; and Trieste in Italy), with multiple possible routes to all of them, considering availability. All datasets were organized and merged to create a unique dataset. Then, each row is organized in such a way that it presents the trips between a pair of control stations, which are the stations that the train crosses between the station of origin and the stations of destination.

Data imputation (i.e., filling missing values using median values for numerical features or the most common class for categorical features) was made to the merged dataset. Then, some additional features were created using available features in the dataset (i.e., feature engineering). Subsequently, categorical encoding was applied to transform the categorical features using dummy variables, while outliers were removed from the numerical features using the interquartile range method and the z-score normalization method was implemented to rescale the numerical features. Finally, a correlation analysis was performed to select the features to train the models to diminish redundant predictors.

The feature to be predicted in this study is the arrival delay time between a pair of control stations. Therefore, the most suitable data-driven approach is regression, which can be considered as supervised ML models used for predicting numeric values. After performing the data pre-processing and feature engineering, we identified a total of 13,894 trips between control stations.

2.2 Data-Driven models and analysis of input features

Data were randomly split into a training and test set (70% and 30%, respectively) with the aim of using the former for training the models and the latter to assess their performance, where each subset was made-up by the target feature to predict (arrival_delay_time) and the remaining independent input features following the same proportion. Then, a suite of ML models that have been extensively and effectively implemented in multiple regression problems was trained to predict the arrival delay time. Furthermore, the relative mean squared error (rMSE) and the coefficient of determination (R2) evaluation metrics were used as loss functions to evaluate data-driven model performance, where the model with a greater value of R2 coefficient, and a lesser value of rMSE, the better.

The model optimization is made using the random search method, with the aim of tuning the parameters of the models and the stratified K-fold cross-validation method was implemented to evaluate the model's performance. After choosing the best ML model for predicting the arrival delay time, the learning curve method was used to identify possible overfitting or underfitting problems.

After obtaining the best data-driven model, the impacts of the features associated with arrival delay time are obtained using the coefficients given by the model itself for every input feature. Coefficients in the output of the models represent the relationship between the given input feature x_i and the target y (i.e., arrival delay time), assuming that all the other features x_j remain constant, which follows the conditional dependence theory. These coefficients indicate the impact of an input feature on the model output, allowing to assess the effect of every single feature on the arrival delay time.

3 RESULTS

3.1 Data-driven models

Several combinations of features were tested in order to identify the combination that achieves better results in the ML models, where the final composition of the dataset is presented in Table 1.

Table 1 - *Composition of the final dataset*

Feature	**Description**	**Feature distribution and statistics**
TARGET: arrival_delay	Arrival delay time [min]	range: -289.0 – 1570.0; median: 12.0, mean: 78.4, std: 245.4
teu_count	Number of TEU (Twenty-foot Equivalent Unit)	range: 0.0 – 98.4; median: 64.5, mean: 59.0, std: 21.0
train_length	Train length [m]	range: 14.0 - 720.0; median: 544.0, mean: 536.5, std: 136.2
total_distance_trip	Distance of the TOTAL trip [km]	range: 84.5 – 1454.1; median: 648.6, mean: 547.5, std: 274.1
departure_delay	Departure delay time [min]	range: -260.0 – 1562.0; median: 20.0, mean: 84.6, std: 237.3
distance_between_control_stations	Distance between control stations [km]	range: 1.5 – 815.3; median: 74.2, mean: 126.6, std: 172.7
weight_per_length_of_train	Train weight over train length [t/m]	range: 0.9 – 4.1; median: 2.2, mean: 2.2, std: 0.5
weight_per_wagon_of_train	Train weight over number of wagons [t/wagon]	range: 19.1 – 119.3; median: 69.5, mean: 67.1, std: 16.5

Then, after implementing a suite of data-driven models using the stratified K-fold cross-validation method, we obtained the initial results presented in Table 2, where the evaluation metrics are presented. Considering that the lightGBM, the CatBoost, and the GBR models have the best performance considering the evaluation metrics, the random search method was implemented to tune the performance of these models, analyzing the behavior of their learning curves to ensure they are correct (i.e., the necessary clear convergence trend between training and cross-validation scores, allowing to foresee if adding more observations to the training of these models will likely improve their performance, decreasing risks of *overfitting* and *underfitting*). Table 2 also presents the overall results, including a comment defining whether the learning curve is good or bad considering the aforementioned behavior.

Table 2 – *Results*

Model		R2	rMSE	Model		R2	rMSE
Initial results.							
Light Gradient Boosting Machine (lightGBM)		0.8579	0.1421	Ridge Regression (ridge)		0.8486	0.1514
CatBoost Regressor (CatBoost)		0.8575	0.1425	Bayesian Ridge (BR)		0.8486	0.1514
Gradient Boosting Regressor (GBR)		0.8566	0.1434	Lasso Regression (lasso)		0.8484	0.1516
Linear Regression (LR)		0.8486	0.1514	Random Forest Regressor (RF)		0.8451	0.1549
Model	**Behavior of the learning curve**	**R2**	**rMSE**	**Model**	**Behavior of the learning curve**	**R2**	**rMSE**
Results of the best models after tuning their parameters using the random search method.							
Initial lightGBM	Good	0.8579	0.1421	Tuned lightGBM	Bad	0.8585	0.1415
Initial CatBoost	Good	0.8575	0.1425	Tuned CatBoost	Bad	0.8584	0.1416
Initial GBR	Good	0.8566	0.1434	Tuned GBR	Bad	0.8581	0.1419

Considering the results presented in Table 2, even though the results are quite similar for some models, the initial lightGBM (an open-source gradient boosting framework) model slightly outperforms for predicting the arrival delay time in freight rail operations considering the evaluation metrics (although any of those are valid options, considering the similarities in the evaluation metrics). Therefore, this model is selected to assess the impact of the input features on the model output, and, furthermore, to analyze the interconnection of disturbances and their subsequent delays.

3.2 Analysis of the input features

Figure 1 provides the importance of each input feature on the magnitude of the initial lightGBM model output, revealing that the departure delay time is the most important feature for predicting the arrival delay time, followed by the distance of the trip (both between control stations and total distance) and the composition of the train (in terms of weight, length and number of wagons).

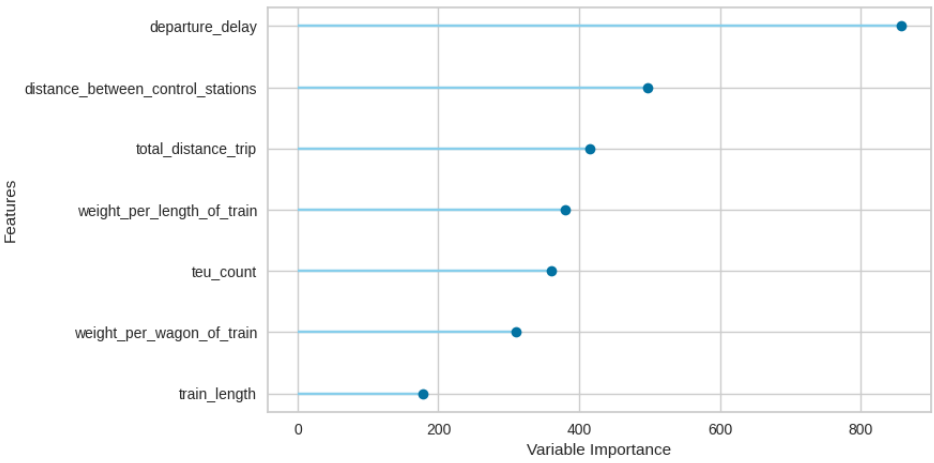

Figure 1 - *Plot showing the input feature importance on arrival delay time resulting from the initial lightGBM model.*

4 DISCUSSION

To the best of our knowledge, this is the first study where the gradient boosting model has been used to predict the arrival delay time in freight rail operation research. Other studies that have obtained acceptable results using different ML models in freight rail operations, are the one made by Pineda-Jaramillo *et al.* (2022) who implemented a CatBoost model to analyze the relationship between the train and the operational features with the disturbances and their subsequent delays, and the studies performed by Wen *et al*. (2019) and Oneto *et al*. (2018), who explored the use of artificial neural networks to predict delays in freight trains.

Pineda-Jaramillo *et al*. (2022) made a similar analysis for freight rail operations, finding that the greater the train weight, length and weight per wagon, and the lower the number of wagons, the greater the probability that the trip will be delayed. Besides, some studies have identified that train length affects the punctuality of passenger and freight trains (Økland and Olsson 2021; Olsson and Haugland 2004; Harris and Godward 1992), whereas Van Der Kooij *et al*. (2017) found that imposing temporary speed restrictions to heavier and longer passenger trains to ensure the safe use of the infrastructure can cause vital delays in the network.

The main contributions of this study are summarized as follows (a) A consistent short-term predictive data-driven model was developed, identifying that the lightGBM implementation of the gradient boosting machine model can predict the arrival delay time in freight rail operations, outperforming other data-driven models; (b) a further analysis of the impacts of the features associated with arrival delay time was made, revealing that the departure delay time, the distance of the trip and the composition of the train are crucial to predict the arrival delay time in freight rail operations; and (c) the short-term prediction model developed in this study can be used as a tool by the National Rail Company of Luxembourg. For instance, through a simple web service, it is possible to know which will be the arrival delay time of a train, and then assess future operational interventions in order to reduce disturbances and their subsequent delays in their freight operations.

References

Barbour, William, Juan Carlos Martinez Mori, Shankara Kuppa, and Daniel B. Work. 2018. "Prediction of Arrival Times of Freight Traffic on US Railroads Using Support Vector Regression." *Transportation Research Part C: Emerging Technologies* 93 (August): 211–227. doi:10.1016/j.trc.2018.05.019.

Bešinović, Nikola, Rob M.P. Goverde, Egidio Quaglietta, and Roberto Roberti. 2016. "An Integrated Micro–Macro Approach to Robust Railway Timetabling." *Transportation Research Part B: Methodological* 87 (May): 14–32. doi:10.1016/j.trb.2016.02.004.

Cacchiani, Valentina, Alberto Caprara, and Paolo Toth. 2010. "Scheduling Extra Freight Trains on Railway Networks." *Transportation Research Part B: Methodological* 44 (2). Elsevier Ltd: 215–231. doi:10.1016/j.trb.2009.07.007.

De Martinis, Valerio, and Francesco Corman. 2018. "Data-Driven Perspectives for Energy Efficient Operations in Railway Systems: Current Practices and Future Opportunities." *Transportation Research Part C: Emerging Technologies* 95 (August). Elsevier: 679–697. doi:10.1016/j.trc.2018.08.008.

Goverde, Rob M.P., Nikola Bešinović, Anne Binder, Valentina Cacchiani, Egidio Quaglietta, Roberto Roberti, and Paolo Toth. 2016. "A Three-Level Framework for Performance-Based Railway Timetabling." *Transportation Research Part C: Emerging Technologies* 67 (June): 62–83. doi:10.1016/j.trc.2016.02.004.

Harris, Nigel, and E.W. Godward. 1992. *Planning Passenger Railways: A Handbook*. Edited by Transport Publishing. UK.

Økland, Andreas, and Nils O. E. Olsson. 2021. "Punctuality Development and Delay Explanation Factors on Norwegian Railways in the Period 2005–2014." *Public Transport* 13 (1): 127–161. doi:10.1007/s12469-020-00236-y.

Olsson, Nils O.E., and Hans Haugland. 2004. "Influencing Factors on Train Punctuality - Results from Some Norwegian Studies." *Transport Policy* 11 (4): 387–397. doi:10.1016/j.tranpol.2004.07.001.

Oneto, Luca, Emanuele Fumeo, Giorgio Clerico, Renzo Canepa, Federico Papa, Carlo Dambra, Nadia Mazzino, and Davide Anguita. 2018. "Train Delay Prediction Systems: A Big Data Analytics Perspective." *Big Data Research* 11 (March). Elsevier Inc.: 54–64. doi:10.1016/j.bdr.2017.05.002.

Pineda-Jaramillo, Juan, Pablo Martínez-Fernández, Ignacio Villalba-Sanchis, Pablo Salvador-Zuriaga, and Ricardo Insa-Franco. 2021. "Predicting the Traction Power of Metropolitan Railway Lines Using Different Machine Learning Models." *International Journal of Rail Transportation* 9 (5). Taylor & Francis: 461–478. doi:10.1080/23248378.2020.1829513.

Pineda-Jaramillo, Juan, William McDonald, Wei Zheng, and Francesco Viti. 2022. "Identifying the Major Causes Associated to Rail Intermodal Operation Disruptions Using Causal Machine Learning." In *101st Transportation Research Board*. Washington, DC.

Van Der Kooij, Rimmert B.K., Andreas D. Landmark, Andreas A. Seim, and Nils O.E. Olsson. 2017. "The Effect of Temporary Speed Restrictions, Analyzed by Using Real Train Traffic Data." *Transportation Research Procedia* 22. Elsevier B.V.: 580–587. doi:10.1016/j.trpro.2017.03.047.

Wen, Chao, Weiwei Mou, Ping Huang, and Zhongcan Li. 2019. "A Predictive Model of Train Delays on a Railway Line." *Journal of Forecasting*, no. February 2019: 470–488. doi:10.1002/for.2639.

Zheng, Wei, and William McDonald. 2021. "Understanding Intermodal Operations Reliability." University of Luxembourg.