Communication

Artificial Intelligence Component of the FERODATA AI Engine to Optimize the Assignment of Rail Freight Locomotive Drivers

Adrian Brezulianu [1,2], Oana Geman [3,*] and Iolanda Valentina Popa [2,4]

1. Faculty of Electronics, Telecommunications and Information Technology, Gheorghe Asachi Technical University of Iasi, 700050 Iasi, Romania; adi.brezulianu@greensoft.com.ro
2. GreenSoft S.R.L., 700137 Iasi, Romania; iolivp@gmail.com
3. Faculty of Electrical Engineering and Computer Sciences, Stefan cel Mare University of Suceava, 720229 Suceava, Romania
4. Faculty of Medicine, "Grigore T. Popa" University of Medicine and Pharmacy, 700115 Iași, Romania
* Correspondence: oana.geman@usm.ro

Abstract: The optimization of locomotive drivers' scheduling in rail freight transportation comes as a necessity for minimizing economic expenses and training investments. The Ferodata AI engine, an artificial intelligence (AI)/machine learning (ML) software module, developed by our team, has integrated a supervised random forest model that automatically assigns conductors to freight transportation orders based on the data about locomotive driver's tiredness score, distance of the driver to the departure point of a transportation order, driver availability, and circulation history. The model proposed by us obtained very good performance metrics on the train set (accuracy: 95%, AUC: 0.9905) and reasonably good and encouraging performance on the test set (accuracy: 84%, AUC: 0.8357). After rigorous testing and validation on external and larger datasets, the automated optimization of locomotive driver assignments could bring operational efficiency, cost savings, regulatory compliance, and improved safety to scheduled rail freight transports.

Keywords: rail freight transportation; machine learning; optimal locomotive driver assignment; artificial intelligence

1. Introduction

A locomotive engineer (driver) is an essential human resource without which a freight train cannot be set in motion [1]. Currently, in Romania, there is still no efficient, automated system to optimize the assignment of engineers to freight trains [2].

Expensive and ample training is provided to all employees in the railway industry, with a particularly rigorous focus on conductors and engineers who operate freight trains [3]. Aspiring engineers are required to undergo comprehensive training that can span over several months, combining intensive classroom instruction, real-world practical training, and regular assessments [4].

The optimization of human resource scheduling in rail freight transportation comes as a necessity for minimizing economic expenses and training investments [5]. However, creating an automated system for engineer assignment is not an easy task.

In the process of allocating locomotive drivers for railway transport, the traffic dispatcher of a railway company must take into account a set of strict legal regulations as well as a set of constraints related to optimizing the use of human resources [6]. Thus, each driver allocated to operate a train on a specific route must meet the requirement regarding the maximum driving/service time, as well as rest time and place. The driving/service times are strictly dependent on the rest time since the last service and the location where the driver claims to have rested. The resting place can be at home or another location if the engineer is in transit.

It should be considered that, in the railway transport process, the company uses a set of dispatch points where trains are formed and where the locomotive drivers' service begins [7]. These dispatch points can be located in the engineer's hometown or elsewhere.

The traffic dispatcher must effectively utilize the locomotive engineer human resource, which is not available in the necessary proportions for railway transport activities, in order to carry out commercial transport operations, while simultaneously complying with all legal requirements and minimizing their overburdening or rest in places other than their domicile [7].

Several papers published so far tackled the subject of automation in railway crew scheduling. One commonly used approach is to employ complete enumeration to generate all possible candidates that satisfy the given constraints and, afterward, apply integer linear programming to solve the problem [8] of crew scheduling in transit systems. Additionally, heuristic methods are widely adopted. In the domain of public transport driver scheduling, a combined approach employing both integer linear programming and heuristics was proposed [9]. Li and Kwan [10,11] proposed a greedy/heuristic approach integrated with a fuzzy genetic algorithm. Frisch et al. [12] introduced a matheuristic approach. Froger et al. [13] presented a solution strategy that combines Lagrangian relaxation with subgradient optimization. Another highly effective method is column generation [14]. A noteworthy study focuses on applying a column generation technique to optimize periodic rail crew schedules [15].

To our knowledge, none of the papers published to date considered methods based on artificial intelligence (AI)/machine learning (ML) for the automatic allocation of locomotive drivers to scheduled transportation orders. Compared to heuristic techniques, ML models could bring the advantage of firstly time efficiency [16], as only the training of the ML models is time-consuming. Once the model is trained, querying the model in practice is performed in no time, while heuristic solutions may be complex and slow. Secondly, they are reliable, as they learn from past experience: the greater the experience, the more efficient the algorithms [17]. On the other hand, heuristic algorithms will not improve their performance over time. They are flexible [18]: AI systems can adapt to changing conditions, such as unexpected delays or new orders by continuously reevaluating and adjusting assignments in real-time. Heuristic approaches may require manual adjustments to accommodate changes. Finally, AI scalability is advantageous: AI solutions can scale to handle many more transportation orders after the model is trained. Heuristic methods may struggle to handle such complexity effectively.

In the current Romanian freight railway system, the allocation of locomotive drivers is carried out manually by designated dispatchers that invest time, great effort, and are exposed to the risk of fatigue and subsequent human error. An AI solution could replace the heavy manual work of the rail dispatchers and solve several of the limitations of the heuristics methods through an automated, time and resource efficient solution for the allocation of locomotive drivers in rail freight transportation.

Moreover, recent special reports from the European Court of Auditors conclude that European freight rail transport is not yet on the right track [19]. Shifting focus from roads and increasing the utilization of alternative freight transportation modes, such as railways, can significantly contribute to making freight transport more environmentally friendly (greener) [19]. There are currently numerous European policies and strategies aimed to transform rail cargo transportation to a more competitive alternative [19]. Our paper aligns with the current European Union effort by proposing an AI/ML method to optimize resource management in rail freight transports.

Our team worked on a software application, the Ferodata system, that captures the technical–dynamic sensory information of the locomotive, based on which it creates the circulation history of the locomotive/train to which it is associated. The circulation history presents, in chronological order, the passing stations, the station stop times, as well as the fuel consumption/electric energy consumption associated with the section of the route. Through the Ferodata mobile application, the locomotive driver reports the rest time and

location at the beginning of each service. These pieces of information are accumulated in the Ferodata system within the performance history. By correlating the performance history with the circulation history, the location of each driver and the route to the current location can be determined. From the circulation history, the route of travel and the required driving times to reach the destination can also be deduced. Thus, the whereabouts of each driver can be known at any given moment, as well as the current driving/service times. Furthermore, it is possible to analyze the working hours situation per workplace, period, and locomotive driver.

We aimed to develop and add to our Ferodata AI engine an innovative AI/ML module that automatically assigns locomotive engineers to freight transportation orders based on collected data such as circulation history and drivers' availability, service time and current location. The module is built to optimize the number of drivers and the amount of service times while complying with the legal regulations. Our final aim is to replace the manual work of rail dispatchers, bringing significant economical and resource management benefits to railway companies and better experience and satisfaction for the final clients.

2. Materials and Methods

2.1. Study Design

A retrospective study was conducted on a sample of 64,768 historical transportation order records. All data were obtained from the database of a Romanian freight railway transportation company. The access to the database was due to our contractual collaboration for the development of the Ferodata software (version 1). More information about the data presented in this study is available on request from the corresponding author.

Data for all transportation orders between the 1 August 2021 and the 30 June 2022 were fetched.

2.2. Data Collection

The following data regarding freight transportation orders were collected for the study from the historical activity of the locomotives: activity start and end dates, start railway station, end railway station, registration number, the assigned driver for that transportation order (the label), location (latitude and longitude) of all the drivers before each transportation order start date, number of working hours of all drivers in the week prior the transportation order start date, and whether at the moment of activity start date the engineers were on medical or annual leave.

The labeling process consisted of several steps. Firstly, all the necessary information about the freight transportation orders (including departure time, departure place, destination) and driver availability (work hours, days off, location with distance to the transportation orders departure place) was collected. Secondly, scheduling started with a blank timetable that represents the available time slots for transportation orders for the next week. The adding of the assignments followed a systematized approach by exhaustive search in the space of all possible assignments that respect legal regulations. The exhaustive search was needed to ensure the optimization of several parameters for the current week: minimization of the number of assigned drivers, maximization of the time spent driving the locomotives (less time spent by the drivers commuting to the place of departure), and fair distribution of work time to all drivers involved. The exact technical details of the manual implementation of driver assignment are confidential data belonging to the railway company and cannot be disclosed in full. Consistency in the labeling process was warranted by weekly quality checks made by appointed control officers and performance evaluators that had to ensure the dispatchers were correctly following the company's protocol for driver assignments. Although this optimization strategy reduces driving personnel expenses, it is also very time expensive and overloads dispatchers. This prompted the need for an automatic solution that can learn from all the labels manually assigned thus far.

2.3. Problem Statement

Our research paper proposes a technical solution that automatically assigns drivers to locomotives for the current work week in a way that optimizes the previously mentioned three parameters: number of assigned drivers, time spent driving the locomotives, and fair distribution of work time to all drivers involved, while complying with the legal regulations.

2.4. Study Protocol

Each record from the study's dataset has three defined columns (two attributes and the outcome variable). The two attributes are: the tiredness score (=the number of working hours of a driver in the week prior to a transportation order start date) and the distance to the departure point (in km) of a scheduled transport. The outcome variable is also a binary variable illustrating whether a driver was assigned to that transportation order or not.

For the current phase of the study, we define the following constraints: a driver should not have more than 40 h of work in a week and should not be on medical or annual leave. Furthermore, we currently assume that driver preferences are not taken into account and there is no enroute change of drivers. In future studies, we will integrate driver preferences and the possibility of enroute change of drivers as parameters for the training of our models.

The records belonging to drivers that are on medical or annual leave at the moment of a scheduled transportation order are excluded from the dataset.

The documented continuous variables were normalized in the range [0, 1].

The initial dataset consisting of 64,768 records was randomly divided into a training set of 51,814 records (80%) and a test set of 12,954 records (20%).

Random forest (RF) and multi-layer perceptron (MLP) classifiers were developed. In RStudio, we utilized the caret::train function to construct the models, taking precautions to prevent overfitting by implementing 10-fold cross-validation. Implementation and mathematical details behind caret models in R are publicly available and can be consulted in the online documentation [20]. To handle imbalanced output classes, we applied the synthetic minority over-sampling technique (SMOTE) alongside the caret::train function. In order to ensure the study's replicability, Table 1 provides an overview of the parameters passed to the caret::train function.

Table 1. Parameters used to train the ML models with the caret::train function.

Method	"rf"/"mlpML"
preProcess	c("scale", "center")
trControl	trainControl(method = "repeatedcv", number = 10, repeats = 10, sampling = "smote")

The sigmoid function was used as the transfer function for the MLP model. The caret::train function automatically tuned three hyperparameters for the mlpML method and one hyperparameter for the RF method. The automatically tuned parameters were the number of neurons in each of the three hidden layers corresponding to the mlpML method design and mtry corresponding to the RF method. The hyperparameters were tuned so as to maximize the model's accuracy. A detailed implementation flow chart of the RF and MLP classifiers is illustrated in Figure 1.

The RF and MLP binary models were developed to predict whether an available locomotive driver should be assigned for a scheduled transportation order considering the current distance to the departure location and the tiredness score. The performance metrics of the RF and MLP models are compared and the algorithm with the highest accuracy is chosen to be integrated into the platform.
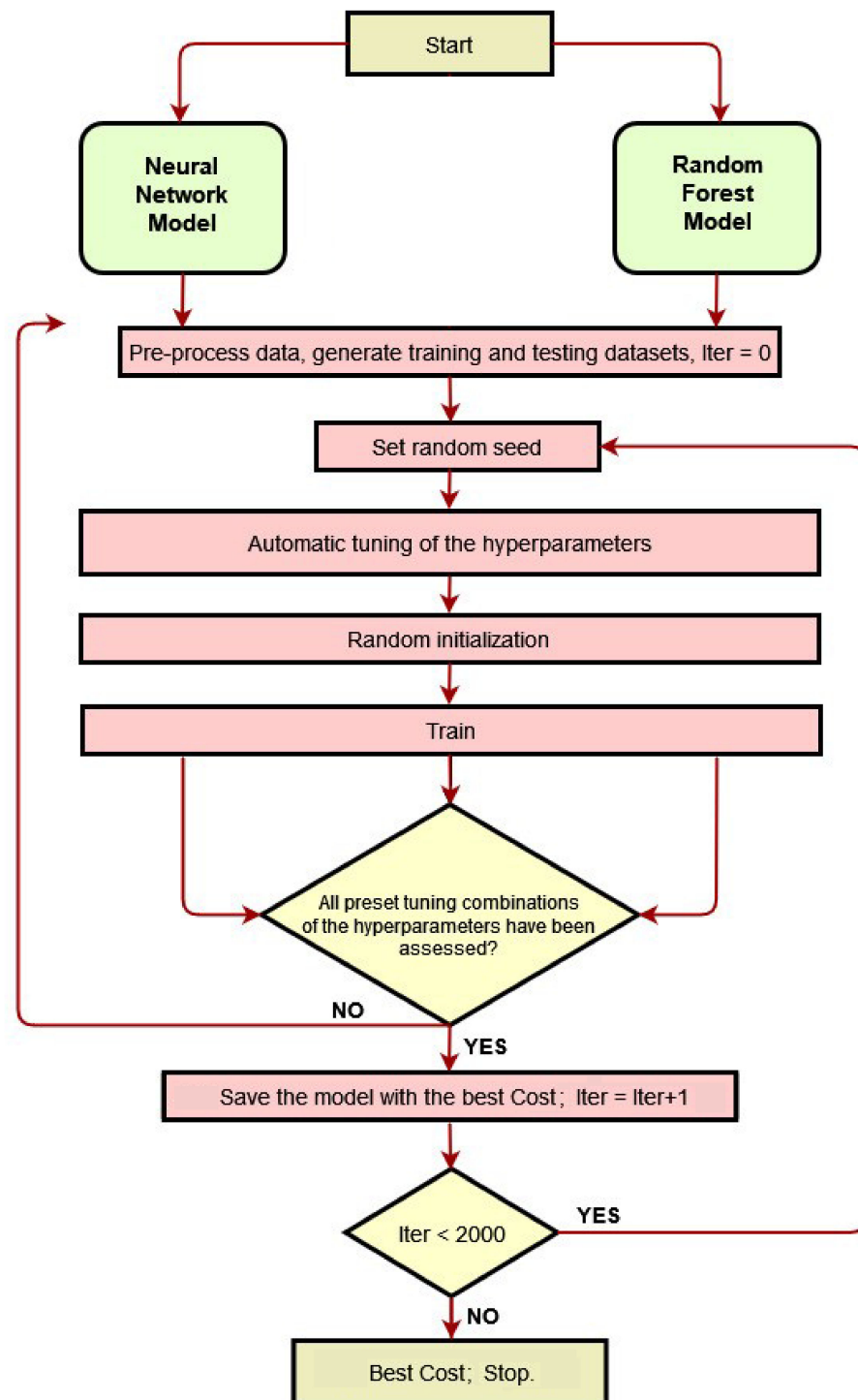
Figure 1. Implementation flow chart of the random forest and multi-layered perceptron models.

Depending on the availability, tiredness score at the moment of the freight transportation order, and distance to the departure point, the driver recommended to the dispatcher for assignment to each transport order is chosen based on the highest probability determined by the chosen ML model.

If a driver already has 40 h of work in the week prior to a scheduled freight expedition or it are on medical or annual leave, they are automatically declared unavailable before querying the ML model.

The preprocessing and models' implementation are performed in RStudio version 2023.03.0+386 (3c53477a, 2023-03-09) for Windows.

Validation of our ML model is a critical step in the development process to ensure that the model can generalize well to new, unseen data. It involves assessing the performance and reliability of the model using data that it has not seen during training. The main goal of validation is to estimate how well the model is expected to perform on new, real-world data. Therefore, we assessed the classification accuracy of the ML models on the test set, as it contains only new, unseen data. We also calculated metrics such as the area under the ROC curve (AUC), sensitivity, specificity, as well as positive predictive values (PPV) and negative predictive values (NPV).

3. Results

The selection of the optimal driver in terms of tiredness score, availability, and distance to the departure railway point was performed according to supervised random forest mathematical optimization in R.

The training of the mathematical optimizer was carried out on 51,814 records and testing on the unseen data was carried out on 12,954 records from the collected Unicom Tranzit data.

For the sake of transparency, we present the first 10 records from the training set in Table 2.

Table 2. The first 10 records from the training set.

Tiredness Score	Distance to Departure Point (km)	Driver Was Selected for the Transport Order (1—'Yes', 0—'No')
0.54	180.94	0
6.53	142.61	0
0	80.5	1
24.17	77.56	0
35.24	70.94	0
33.53	96.175	0
26.08	27.18	1
23.16	98.37	0
17.46	151.76	0
19.93	77.14	0

A summary of the continuous parameters of the whole dataset is presented in Table 3. Additionally, we present the information that summarizes the train set and the test set in Tables 4 and 5, respectively. No significant differences were found between the training and test sets regarding each of the collected parameters (continuous or categorical). Numerical data are presented as median with interquartile range, due to the non-normal distribution. Categorical data are described by number of occurrences and percentage.

After the training step, we obtained the RF and MLP models that we tested on both the training set and the test set (unseen data).

Table 3. Summary of the parameters of the whole dataset.

	Tiredness Score	Distance to Departure Point	Driver Was Selected for the Transport Order
Min	0	0	
1st interquartile range	0	65.75	
Median	6.28	105.15	12,933 (19.97%)
3rd interquartile range	18.6	152.35	
Max	61	313.85	

Table 4. Summary of the parameters of the training set.

	Tiredness Score	Distance to Departure Point	Driver Was Selected for the Transport Order
Min	0	0	
1st interquartile range	0	49.98	
Median	5.51	120.84	10,412 (20.09%)
3rd interquartile range	20.3	159.12	
Max	59.8	313.85	

Table 5. Summary of the parameters of the test set.

	Tiredness Score	Distance to Departure Point	Driver Was Selected for the Transport Order
Min	0	0	
1st interquartile range	0	71.51	
Median	7.02	98.32	2571 (19.85%)
3rd interquartile range	16.98	149.76	
Max	61	310.2	

After obtaining the RF model, the request to display it in Rstudio results in the following Rstudio console response, which demonstrates the main parameters and training performance of the model (10-fold cross validation, 10 repeats, a mtry value of 2, accuracy of 0.8880513, and a kappa value of 0.6320161):

```
Random Forest
Pre-processing:  scaled (2), centered (2)
Resampling:  Cross-Validated (10 fold, repeated 10 times)
Additional sampling using SMOTE prior to pre-processing

Resampling results:

Accuracy Kappa
0.8880513 0.6320161

Tuning parameter 'mtry' was held constant at a value of 2
```

The performance metrics on the training and test sets for the RF model are illustrated in Table 6.

Table 6. The performance metrics of the RF model on the training and test sets.

	Random Forest	
	Train Set	Test Set
Accuracy	99%	88%
95% CI	(0.9847, 0.9932)	(0.8549, 0.9079)
p value	<0.001	0.001
Sensitivity	99%	92%
Specificity	99%	72%
PPV	100%	94%
NPV	95%	63%
AUC	0.9927	0.8998

The ROC curves of the RF classifier for both the training and test sets can be visualized in Figure 2.
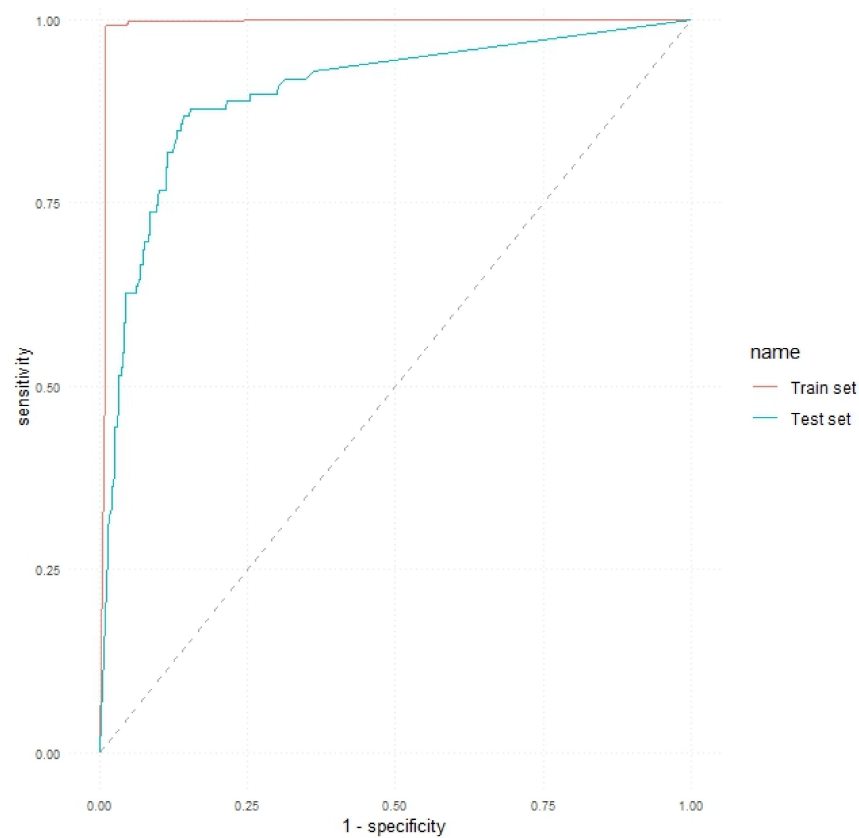

Figure 2. ROC curves obtained by the RF model on the training and test sets.

After obtaining the MLP model, we displayed it in the RStudio console, in a similar manner as we did with the RF classifier, and we obtained the following result. This result also demonstrates the main parameters and training performance of the model (10-fold cross validation, 10 repeats, preprocessing modes, one hidden layer containing five neurons, and accuracy and kappa values for the different numbers of neurons in the hidden layer):

```
Multi-Layer Perceptron, with multiple layers

2 predictor
2 classes:  '0', '1'

Pre-processing:  scaled (2), centered (2)
Resampling:  Cross-Validated (10 fold, repeated 10 times)
Additional sampling using SMOTE prior to pre-processing

Resampling results across tuning parameters:

layer1 Accuracy Kappa
1 0.8107440 0.4717333
3 0.8100957 0.4679649
5 0.8390885 0.5161155

Tuning parameter 'layer2' was held constant at a value of 0
Tuning parameter 'layer3' was held constant at a value of 0
 Accuracy was used to select the optimal model using the
largest value.
```

```
    The final values used for the model were layer1 = 5,
layer2 = 0 and layer3 = 0.
```

The performance metrics on the training and test sets for the MLP model are illustrated in Table 7.

Table 7. The performance metrics of the MLP model on the training and test sets.

	Multi-Layer Perceptron	
	Train Set	**Test Set**
Accuracy	85%	86%
95% CI	(0.8346, 0.8637)	(0.8278, 0.8852)
p value	0.045	0.012
Sensitivity	86%	89%
Specificity	78%	73%
PPV	95%	94%
NPV	53%	58%
AUC	0.8994	0.8857

The ROC curves of the MLP classifier for both the training and test sets are illustrated in Figure 3.
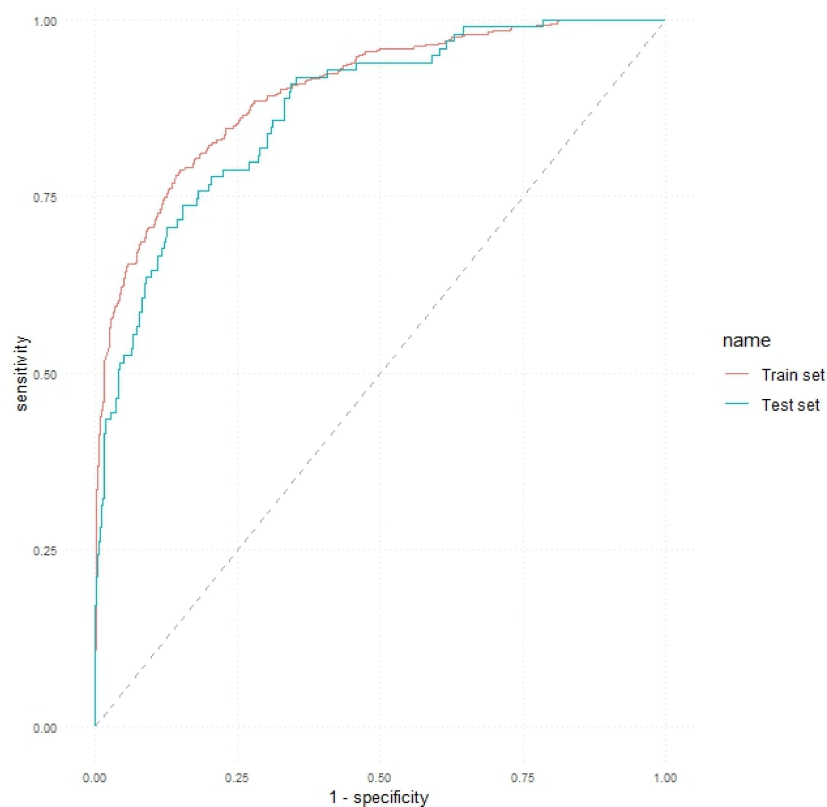
Figure 3. ROC curves obtained by the MLP model on the training and test sets.

As can be seen from the comparison of the RF versus MLP models, the RF classifier achieved higher performance metrics (accuracy, AUC) on both the training and test sets. Therefore, the RF model was chosen to be integrated into the platform.

For each transport order, the driver that is available and has the highest probability of being chosen by the RF model, is selected to be recommended to the dispatcher for assignation to the transport order.

In order to exemplify the practical use of our model, we present in Table 8 two real-world use cases/scenarios.

Table 8. Illustration of two real-world use cases.

Scenarios	Date of Scheduled Transportation Order	Departure Point	Destination Point	The Driver Manually Assigned by the Dispatcher	The Driver Recommended by Our Platform
Scenario 1	1 June 2022, 00:36	PLOIESTI VEST	BASARABI	Anonymized driver 4	Anonymized driver 4
	Result: The platform decision matched the human decision.				
Scenario 2	31 May 2022, 17:14	GHERLA	AIUD	Anonymized driver 12	Anonymized driver 12
	Result: The platform decision matched the human decision.				

For the transportation order scheduled on 1 June 2022 at 00:36, destined to depart from PLOIESTI VEST to BASARABI, the dispatcher manually assigned as driver: anonymized driver 4. The platform was run, and it also recommended as driver: anonymized driver 4. Therefore, the platform decision matched the human decision. This proves that we achieved our goal to develop an automatic module that could replace the manual work of assigning drivers to a locomotive, that is nowadays performed by the dispatcher.

4. Discussion

The Ferodata AI optimization module retrieves data from the database, performs a forecast regarding the most suitable locomotive driver to be allocated for expeditions considering the location of the driver, rest time, and availability. It should be considered that, as new data are gathered, the model will undergo retraining and retesting in order to improve its performance.

The locomotive driver serves as a vital link within the rail transport chain and holds an integral role in its operations [21]. Nevertheless, this profession has encountered a range of challenges in recent years [22]. Factors such as a persistent shortage of qualified drivers, mounting stress levels, and deteriorating working conditions have arisen [23]. As a result, there exist substantial opportunities for innovating the organization of this job position [24].

Our study aimed to propose an effective and innovative method to improve the management of this job position, in the current context. To our knowledge, there has been no scientific paper published to date that proposed a random forest model for automatic allocation of drivers on rail freight transportation orders.

The model proposed by us obtained excellent performance metrics on the train set (accuracy: 99%, AUC: 0.9927) and very good and encouraging performance metrics on the test set (accuracy: 88%, AUC: 0.8998). With the increase in the size of the database and access to more variables (driver skills, driver preferences, enroute change of drivers, etc.), the model will be retrained and improved.

The ML model proposed by us provides several valuable contributions to the logistics of rail freight transportation industry: optimized resource allocation, reduced human error, cost savings (minimize labor cost), real-time adaptability, scalability, and ensuring compliance with industry regulations, safety standards, and labor laws.

Consistent scientific evidence shows that the primary outcome of economic significance would be the reduction in costs through the efficient utilization of train resources, particularly train personnel, including vehicle drivers [25]. Efficient driver assignment through automation can lead to cost savings by reducing overtime expenses, optimizing resource utilization, and minimizing disruptions due to unplanned driver unavailability [26].

It also allows for better planning of driver schedules and potential reduction in the need for additional staff [27].

Automation allows for the optimization of driver assignments based on various factors such as availability, and rest requirements. This leads to more efficient utilization of resources and reduces idle time or unnecessary delays in the transport process [28]. Automated optimization helps in maximizing productivity by ensuring that drivers are consistently engaged in transporting freight. This results in improved overall efficiency, better resource allocation and a balanced workload distribution in the rail freight operations, preventing underutilization or overburdening of drivers.

The planning and execution processes can be simplified by handling the complex task of driver assignment. It eliminates manual or ad-hoc approaches, enabling faster decision-making and reducing administrative burdens. The system can generate optimized driver schedules, taking into account factors such as transport requirements, driver availability, and regulatory constraints. The system could easily be configured to consider and enforce regulatory requirements such as maximum driving times, rest periods, and other labor regulations [29]. By automating the assignment process, it helps ensure compliance with these regulations, minimizing the risk of violations and associated penalties.

Regarding safety, the system can also help manage driver fatigue, ensuring that drivers are adequately rested and capable of safely operating the locomotives. This contributes to maintaining a high level of safety in rail freight operations.

Real-time adaptability is another advantage of using an automated optimization system [30]. When disruptions occur, such as driver unavailability due to illness or unexpected events, automated optimization systems can quickly analyze the situation and reassign available drivers to fill the gaps [31]. By identifying the most suitable replacements based on various factors like tiredness score, proximity, and regulatory compliance, the system minimizes the impact of disruptions on the overall schedule. In the event of disruptions along a specific route or at a particular location, the automated optimization system can rapidly reallocate an optimal driver for the alternative routing by analyzing available drivers and resources, ensuring timely delivery, and minimizing the impact of disruptions on the entire transport network.

Limitations

Firstly, one of the primary limitations of our study is the degree of data availability for the automated optimization. Access to comprehensive and accurate data on company's key performance indicators, driver skills, driver preferences, enroute change of drivers, etc. would allow for a more complex analysis and the more realistic modeling of real-life scenarios. One of our future prospects is to integrate such parameters (when available) in the training or testing of our models. Secondly, our dataset size is rather small, making us uncertain on the model's behavior on large-scale optimization problems. Thirdly, the testing of the models was carried out on internal data. Scalability and computational efficiency are yet to be ensured and evaluated on larger and external datasets, which is another future prospect of our research.

5. Conclusions

We presented the Ferodata AI engine, an AI/ML optimization module, developed by our team, that automatically assigns drivers to freight transportation orders based on the data about drivers, workplaces, driving/service times, and circulation history. The model proposed by us is the first of its kind to be published and obtained excellent performance metrics on the train set and very good and encouraging performance metrics on the test set. The presentation of the Ferodata AI engine and the publication of this research paper for the rail freight transportation industry can have several important implications and benefits for the industry as a whole. One of the key benefits of an AI/ML optimization module like the Ferodata AI engine is the potential for significant efficiency improvements. By automatically assigning drivers to freight transportation orders based on various data points, idle time

can be reduced and overall resource allocation can be improved. This leads to cost savings for transportation companies and potentially lower prices for customers. The ability to assign drivers to orders with precision can lead to better service quality. Timely deliveries and reduced delays can enhance customer satisfaction, making rail freight transportation a more attractive option for businesses that rely on it. AI-driven optimization can help rail freight companies make the most of their resources. It can balance workloads, prevent the overworking of certain drivers, and ensure that resources are used efficiently across the network. Optimized driver assignments can also lead to safety improvements. By managing driver schedules and driving times more effectively, it is possible to reduce fatigue-related accidents and enhance overall safety within the industry. By optimizing driver schedules and ensuring compliance with regulations regarding working hours and rest breaks, the Ferodata AI engine can help rail freight companies maintain regulatory compliance and avoid fines or penalties. After rigorous testing and validation on external and larger datasets, the automated optimization of locomotive driver assignments could bring operational efficiency, cost savings, regulatory compliance, and improved safety to scheduled rail freight transports. If the AI/ML optimization module proves successful, it may encourage other rail freight companies to adopt similar technologies. This can lead to an industry-wide shift towards greater efficiency and improved services.

References

1. Fan, C.; Huang, S.; Lin, S.; Xu, D.; Peng, Y.; Yi, S. Types, Risk Factors, Consequences, and Detection Methods of Train Driver Fatigue and Distraction. *Comput. Intell. Neurosci.* **2022**, *2022*, 8328077. [CrossRef] [PubMed]
2. Insights into the Railway Infrastructure in Romania. Available online: https://china-cee.eu/2023/04/28/romania-social-briefing-insights-into-the-railway-infrastructure-in-romania/ (accessed on 22 May 2023).
3. Europe's Rail Joint Undertaking Multi-Annual Work Programme. Available online: https://rail-research.europa.eu/wp-content/uploads/2021/12/20211222_mawp_v1_agreed-in-principle_clean.pdf (accessed on 25 May 2023).
4. Abbas, Y.; Martinetti, A.; Nizamis, K.; Spoolder, S.; van Dongen, L.A.M. Ready, trainer . . . one*! discovering the entanglement of adaptive learning with virtual reality in industrial training: A case study. *Interact. Learn. Environ.* **2021**, *31*, 3698–3727. [CrossRef]
5. Zhang, X.; Li, L.; Zhang, J. An optimal service model for rail freight transportation: Pricing, planning, and emission reducing. *J. Clean. Prod.* **2019**, *218*, 565–574. [CrossRef]
6. Campos, J.; Cantos, P. Rail Transport Regulation. Available online: https://documents1.worldbank.org/curated/en/284281468764045820/135535322_20041117180643/additional/multi-page.pdf (accessed on 22 May 2023).
7. Janota, A.; Pirník, R.; Ždánsky, J.; Nagy, P. Human Factor Analysis of the Railway Traffic Operators. *Machines* **2022**, *10*, 820. [CrossRef]
8. Han, A.F.; Li, E.C. A constraint programming-based approach to the crew scheduling problem of the Taipei mass rapid transit system. *Ann. Oper. Res.* **2014**, *223*, 173–193. [CrossRef]
9. Fores, S.; Proll, L.; Wren, A. TRACS II: A hybrid IP/heuristic driver scheduling system for public transport. *J. Oper. Res. Soc.* **2002**, *53*, 1093–1100. [CrossRef]

10. Li, J. A Self-Adjusting Algorithm for Driver Scheduling. *J. Heuristics* **2005**, *11*, 351–367. [CrossRef]
11. Li, J.; Kwan, R.S.K. A fuzzy genetic algorithm for driver scheduling. *Eur. J. Oper. Res.* **2003**, *147*, 334–344. [CrossRef]
12. Frisch, S.; Hungerländer, P.; Jellen, A. On a Real-World Railway Crew Scheduling Problem. *Transp. Res. Procedia* **2022**, *62*, 824–831. [CrossRef]
13. Froger, A.; Guyon, O.; Eric, P. *A Set Packing Approach for Scheduling Passenger Train Drivers: The French Experience*; HAL: Lyon, France, 2015.
14. Heil, J. A Solution Approach for Railway Crew Scheduling with Attendance Rates for Multiple Networks. In Proceedings of the Operations Research Proceedings 2018: Selected Papers of the Annual International Conference of the German Operations Research Society (GOR), Brussels, Belgium, 12–14 September 2018; Springer: Cham, Switzerland, 2019; pp. 547–553.
15. Janacek, J.; Kohani, M.; Koniorczyk, M.; Marton, P. Optimization of periodic crew schedules with application of column generation method. *Transp. Res. Part C Emerg. Technol.* **2017**, *83*, 165–178. [CrossRef]
16. Zongheng, Y. *Machine Learning for Query Optimization*; University of California: Berkeley, CA, USA, 2022.
17. Zhang, C.; Wu, Y.; Ma, Y.; Song, W.; Le, Z.; Cao, Z.; Zhang, J. A review on learning to solve combinatorial optimisation problems in manufacturing. *IET Collab. Intell. Manuf.* **2023**, *5*, e12072. [CrossRef]
18. Wang, H.; Jacob, V.; Rolland, E. Design of efficient hybrid neural networks for flexible flow shop scheduling. *Expert Syst.* **2003**, *20*, 208–231. [CrossRef]
19. ECA. EU Freight Transport: The Truck Continues to Rule. Available online: https://www.eca.europa.eu/Lists/News/NEWS202303_27/INSR-2023-08_EN.pdf (accessed on 5 September 2023).
20. The Caret Package. Available online: https://topepo.github.io/caret/ (accessed on 29 September 2023).
21. Wangai, A.W.; Rohacs, D.; Boros, A. Supporting the Sustainable Development of Railway Transport in Developing Countries. *Sustainability* **2020**, *12*, 3572. [CrossRef]
22. Wang, Z.; Liu, X. Cyber security of railway cyber-physical system (CPS)—A risk management methodology. *Commun. Transp. Res.* **2022**, *2*, 100078. [CrossRef]
23. Hranický, M.P.; Černá, L.; Kuka, A. Application of the Theory of Constraints for Railway Personnel. *Transp. Res. Procedia* **2021**, *53*, 31–38. [CrossRef]
24. Naweed, A.; Chapman, J.; Allan, M.; Trigg, J. It Comes with the Job: Work Organizational, Job Design, and Self-Regulatory Barriers to Improving the Health Status of Train Drivers. *J. Occup. Environ. Med.* **2017**, *59*, 264–273. [CrossRef]
25. Vojtek, M.; Kendra, M.; Zitrický, V.; Široký, J. Mathematical approaches for improving the efficiency of railway transport. *Open Eng.* **2020**, *10*, 57–63. [CrossRef]
26. Andreasson, R.; Jansson, A.A.; Lindblom, J. The coordination between train traffic controllers and train drivers: A distributed cognition perspective on railway. *Cogn. Technol. Work* **2019**, *21*, 417–443. [CrossRef]
27. Union Internationale des Chemins de Fer. Final Report Summary—CAPACITY4RAIL (Increasing Capacity 4 Rail Networks through Enhanced Infrastructure and Optimised Operations). Available online: https://cordis.europa.eu/project/id/605650/reporting/it (accessed on 22 May 2023).
28. Khmeleva, E.; Hopgood, A.A.; Tipi, L.; Shahidan, M. Fuzzy-Logic Controlled Genetic Algorithm for the Rail-Freight Crew-Scheduling Problem. *KI—Künstliche Intell.* **2018**, *32*, 61–75. [CrossRef]
29. Demiridis, N.; Pyrgidis, C. Getting freight trains back on track–How railway undertakings, infrastructure owners and regulators can navigate the main dilemmas in freight business to drive sustainable growth. *Front. Sustain.* **2022**, *3*, 903945. [CrossRef]
30. Veelenturf, L.P.; Potthoff, D.; Huisman, D.; Kroon, L.G.; Maróti, G.; Wagelmans, A.P.M. A Quasi-Robust Optimization Approach for Crew Rescheduling. *Transp. Sci.* **2016**, *50*, 204–215. [CrossRef]
31. Brandenburger, N.; Jipp, M. Effects of expertise for automatic train operations. *Cogn. Technol. Work* **2017**, *19*, 699–709. [CrossRef]