Hamburg University of Technology

Machine Learning in Motion

Forecasting of energy of freight trains using machine learning

Ing. Andrew Agib, M.Sc. Marvin Kastner,
Prof. Dr.-Ing. Carlos Jahn

Abstract

Maritime logistics plays a big role in world trade. A lot of factors affect the cost efficiency of this complex process. For hinterland transportation often freight trains are used. This project targets at forecasting the energy consumption of such a freight train specially at peak time windows. This allows to optimize train schedule to avoid peaks in the energy consumption which in turn will reduce the electricity bill.

Introduction

The energy consumption of freight trains is influenced by many different factors. Mistakes in calculating the energy demand leads to peaks of energy consumption, which can result in additional costs of more than 50.000€. In this project, the main points of interest are the energy consumed and recuperated. These two values are influenced by the length, weight, and speed of the train as well as the inclination and further characteristics of the rail tracks. The data used was collected by BoxXpress.de GmbH, a freight trains company that moves more than 670,000 TEU on the rail network for customers and not only relieve congested roads but also the environment.

Methods and implementation

First, data cleaning was of major importance:

- Verification of the energy consumption of trains
- Removing irrelevant attributes
- Then the next point was to merge the timely data from different data sources. One dataset held the information about the rail tracks, one dataset held the information about when a train passed which checkpoint, and the last dataset held the information about in which time window how much energy was consumed and recuperated.
- The correlation "used to get relation between parameters" was:

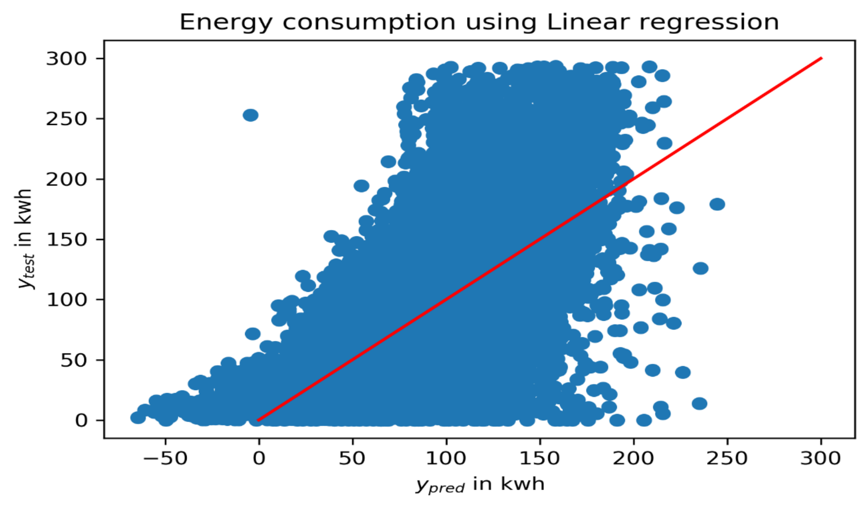
	Speed	Inclination	Distance	Weight	Length	TEU (%)	Energy	Recuperated
Speed	1	-0.02351	0.052046	-0.00018	-0.00088	-0.00266	-0.09456	-0.08652
Inclination	-0.02351	1	0.053427	0.029378	-0.00885	-0.01869	0.115483	0.105686
Distance	0.052046	0.053427	1	0.005385	-0.03916	0.003175	0.24606	0.102697
Weight	-0.00018	0.029378	0.005385	1	0.136421	0.395261	0.03804	0.061765
Length	-0.00088	-0.00885	-0.03916	0.136421	1	0.1178	-0.00784	-0.00688
TEU (%)	-0.00266	-0.01869	0.003175	0.395261	0.1178	1	0.026364	0.02737
Energy	-0.09456	0.115483	0.24606	0.03804	-0.00784	0.026364	1	0.696871
Recuperated	-0.08652	0.105686	0.102697	0.061765	-0.00688	0.02737	0.696871	1

- Then the IQR method "method used to remove outliers from the system" from 25% to 75%, was applied.

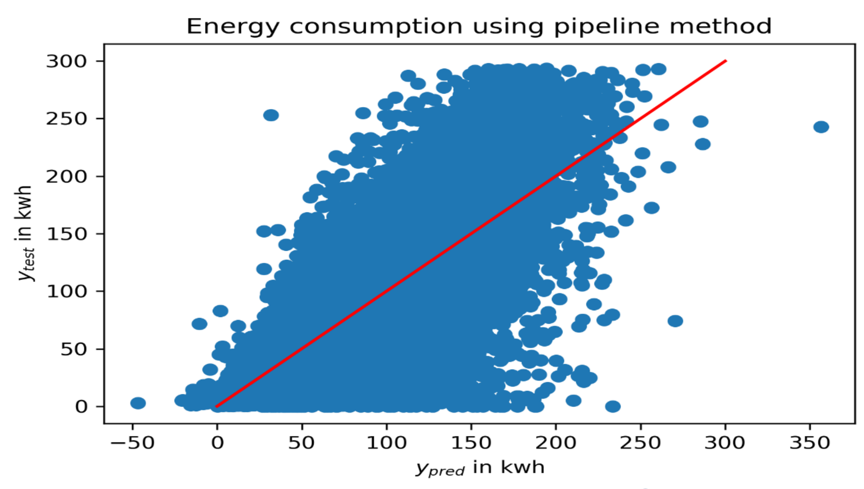
After that machine learning techniques were used. Decision trees, and pipelines were used in a k-fold validation setting. MLP neural network, and Random forest were implemented; with removing the peaks only of the energy, as IQR showed bad results with them.

R^2"used to get how accurate predictions, compared to the actual" was used to validate the accuracy of diffrent systems.

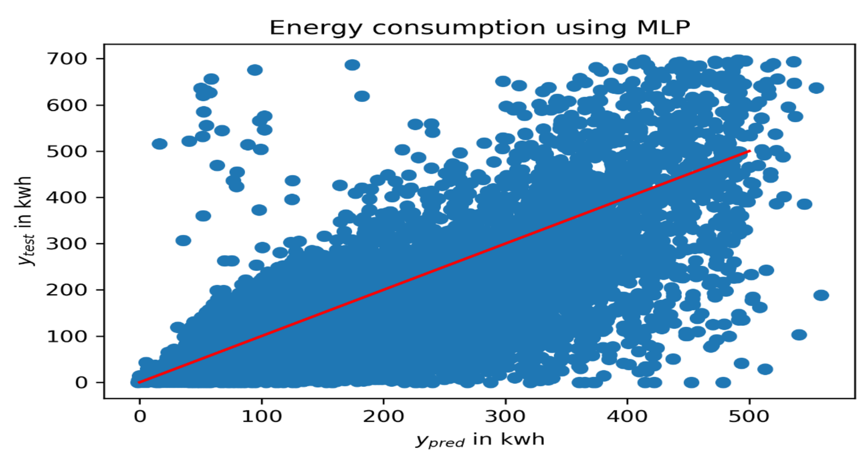
- The decision tree showed an R^2 of 0.31



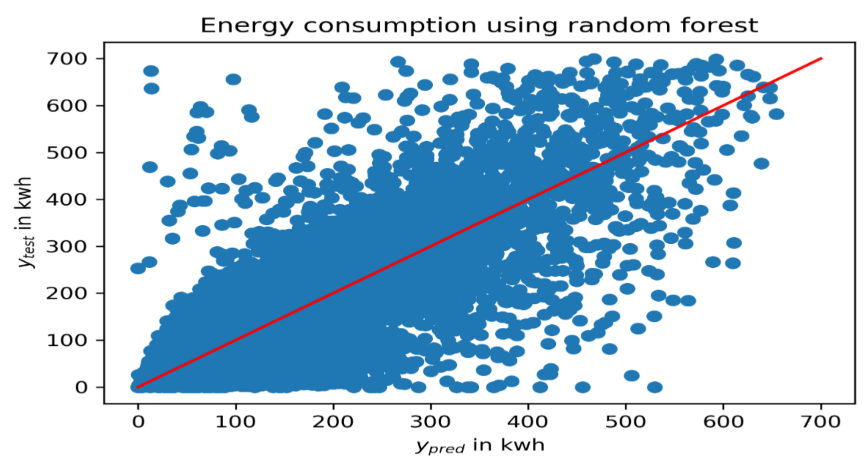
- The Pipeline showed an R^2 of 0.55



- The MLP regression showed an R^2 of 0.67



- The Random forest showed an R^2 of 0.7



Conclusion and further work

After comparing the four techniques, it was clear the most accurate one was Random forest. Most likely this is because a multitude of decision trees is capable of reflecting the complexity of the underlying processes of energy consumption in the best way.

For future work that could be implemented based on the results of this project, more details about the rail tracks can be incorporated. Second, the predictions can be used for an automated recommendation system to avoid peaks for the operation of a whole freight train fleet.

The presented work has been executed in the scope of „MaLiTuP – Machine Learning in Theory and Practice", funded by the Federal Ministry of Education and Research (BMBF) (project number: 01IS17077)

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung