NORDIC MACHINE INTELLIGENCE

**Nordic Machine Intelligence, vol. 10, pp. 1–3, 2021**
Received 1 Jan 2021 / published 21 Jul 2021
https://doi.org/10.2478/nmi-some-doi

# Team NoDig - Decision Support for Sustainable Worksites

Lars Martin Hodne[1 2], Ole-Magnus Høiback[1], Mats Stensrud Skui[1], Øystein Stavnes Sletta[1], Andreas Thyholt Henriksen[1]

1. *Norconsult Digital, Trondheim*
2. *E-mail any correspondence to: lars.martin.hodne@norconsult.com*

**Abstract**
*RoadAI: Reducing emissions in road construction*[1] is a competition hosted by the Norwegian Artificial Intelligence Research Consortium (NORA) in collaboration with Skanska, Sintef, and other partners. Data from a Skanska construction site has been given to all participants. This report describes an accurate automatic load and dump cycle detector and a daily report generator that can provide real-time insight into mass movement, idle machines, and machine-specific statistics. The transparency, deployment, and further development of the presented solutions are discussed.

**Keywords:** RoadAI; machine learning; LightGBM; agglomerative clustering; automated decision support

## Introduction

Road construction plays a pivotal role in societal development, and facilitating modern life. However, it carries a significant environmental impact, and the construction industry has been said to lag behind other industries in innovation [2]. The RoadAI competition aims to leverage data for emission reduction in construction projects.

## Materials and methods

The RoadAI dataset includes GPS data, vibration data, AEMP machine data, and drone photos. The delivered work includes two independent algorithms, each catering to one task. The algorithms use GPS data, with interactive maps using drone imagery as overlays. For more detail on the data, see RoadAI [1].

The contributions are presented in two parts: a daily report entailing mass transfer, idle time, machine-specific statistics, highlighted areas with ongoing load or dump activity, and the Load And Dump (LAD) algorithm that detects dump and load cycles.

### *Daily Report*

The daily report presents regions with mass movement and heatmaps of regions with idle machines.

#### Map of Mass Movement

An interactive map displays daily load and dump zones, alongside information on the total tonnage moved for each material and the most productive machines, measured in $t/hr$. Scree plot statistics are used to select the initial number of zones. However, the user can select the number of zones manually, and filter between trucks and dumpers for more granular analysis. The load and dump zones are ranked by mass moved. Agglomerative clustering paired with the convex hull algorithm is used to calculate polygons for each zone on the map. The drone images that are closest to the selected date are overlaid on the map for Skaret and Nordlandsdalen.

#### Idle Time

Idle time is reported with an aggregated timeline of idle machines and what action (load or dump) is expected next. The positions of idle machines can be visualized on the map with a heat map overlay during peak hours.

To account for GPS inaccuracy, a machine is considered "idle" if it is traveling less than 5 km/h. While this definition will count load and dump as idle time, this inaccuracy should be close to constant when averaged over the number of trips. Therefore, a baseline can be calculated to offset the slightly high idle time measure, e.g., the median idle time for the past week, which, for the first 5 days of the GPS data, was approximately 6 minutes for trucks moving stone. Comparing the baseline idle time with overall idle time could give insights into excess emissions.

### *Automatic Load and Dump Detection with Light Gradient-Boosting Machine*

The LAD algorithm is based on *Light Gradient-Boosting Machine* (LightGBM). LightGBM is an ensemble framework that uses decision trees as weak learners and can tackle regression, ranking, and classification tasks. LightGBM improves efficiency with *Gradient-based One-Side Sampling* (GOSS) and *Exclusive Feature Bundling* (EFB). These innovative features significantly reduce training time by discarding less important data, while still delivering state-of-the-art accuracy [3].

An augmented data set was made from the GPS data, which includes features such as speed, acceleration, and position. These variables capture shifts in latitude and longitude as well as directional velocities in both the north/south and east/west axes.

For a given machine on a given day, the GPS data is divided with an 80/20 training and test split to ensure that no machine is overrepresented in either split. By aggregating data for the selected machine type, the model is trained to produce predictions on all machines for any included day.

The model offers two modes of prediction: one based on aggregated timestamps and another on discrete data points. While the former approach typically delivers higher predictive accuracy, it sacrifices the resolution of the output compared to the discrete mode. This makes it less suitable for accurately identifying the precise timing of specific operational events, such as driving, dumping, or loading activities.

The model utilizes the default hyperparameter of the LightGBMClassifier[1] with an early stopping criterion in combination with a large upper limit on the number of iterations. This approach aims to minimize the risk of overfitting. For further use, the authors recommend performing an automatic hyperparameter tuning process for a better fit to the data.

## Results

The daily report tool is available as a jupyter notebook in the delivered code (see `daily_report_demo.ipynb`[2]).

The classification results for automatic load and dump prediction with LightGBM are listed in Table 1.

| Class | Precision | Recall | F1-score |
|---|---|---|---|
| Drive | 0.997 | 0.996 | 0.996 |
| Load | 0.839 | 0.887 | 0.862 |
| Dump | 0.845 | 0.865 | 0.855 |

Table 1: Summary statistics of the classification model. Five timestamps were consolidated into a single data point for this analysis.

## Discussion

The LAD algorithm demonstrates the potential to automate load and dump registration for drivers. Automatic load and dump registration should increase the efficiency of drivers, and equally important increase overall data quality by minimizing human error and enabling further analysis. The daily report can also equip decision-makers with insight to uncover operational inefficiencies and increase sustainability.

The potential benefits of having more continuous vibration data should not be understated. The proposed algorithms could be improved further by identifying when a machine is idle, if its engine is running, or if it is shaken during loading and dumping. This, paired with GPS data, could create high-importance features for the LAD algorithm.

### *Transparency*

Machine learning algorithms often lack the transparency needed to understand how a conclusion was reached. Conversely, LightGBM provides feature importance values that explain its reasoning, and the model is fed with covariates depending on GPS data only. Therefore, the algorithm should not introduce ethical issues with transparency any more than the current manual implementation.

The daily report reveals information about productivity and idle time on the machine level. Assuming that it is possible to match drivers with machine IDs, it could be possible to construct performance measures that rank drivers. Consequently, restrictions should be made on how the tool is used to analyze individual drivers to follow enterprise ethical guidelines. Conversely, the use of heatmaps to show idle positions can alleviate concerns about displaying the exact positions of drivers.

### *Deployment*

Automatic detection of load and dump cycles can be deployed to a server where it can access datastreams, and store predictions with GPS data. Seeing that drivers will still need to report load and quantity with LAD, one can have a phase of both automatic and manual registration to fine-tune the algorithm. Hence, deployment should not introduce new procedures for the drivers.

The daily report is meant as an insight tool for decision support. It could be deployed to any application monitoring or business intelligence service, where it would be accessible to decision-makers such as construction administrators. One could also develop a standalone application for the purpose.

### *Limitations*

The vibration data is missing data between the dump and the start of the next trip. It could improve the performance of the load and dump detection algorithm if this is addressed. The machine (AEMP) data is not used since the latitudes and longitudes referenced for each machine were not found to correlate to the GPS data.

### *Conflict of Interest*

The authors state no conflict of interest.

[1] lightgbm.LGBMClassifier

[2] NoDig GitHub repository

## References

1. Riemer-Sørensen S, Holmestad HMB, Lars Horn KM, Singstad BJ, Sabry Razick JCD, and Hansen BMT. RoadAI: Reducing emissions in road construction. Nordic Machine Intelligence 2023; .:1–4. DOI: 10.5617/nmi.10315
2. Ibrahim M and Moselhi O. Automated productivity assessment of earthmoving operations. J. Inf. Technol. Constr. 2014; 19:169–84. Available from: https://api.semanticscholar.org/CorpusID:73724135
3. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, and Liu TY. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems 2017; 30:3146–54