A Comparative Analysis of AI Models for Transit Time Prediction in Transportation Systems

Tuan Vu[1], Ali Jedari Heidarzadeh[1], Seyedamir Ahmadi[1], Wael M. Mohammed[1], Mika Tuomola[2], Jose Luis Martinez Lastra[1]

[1] *FAST-Lab, Faculty of Engineering and Natural Sciences, Tampere University, Tampere, Finland*
[2] *GMM Finland Oy, Santastentie 197, 38950 Honkajoki, Finland*

Abstract

The Time-critical term emphasizes the dominance of time as a factor in systems, operations, processes, and activities. As an example, processing animal by-products is a time-critical process as the material can rapidly degrade and become potentially harmful, hence, not suitable for using as raw material in added value products. In this industry, time estimation and prediction enlarge the margin for making decisions in logistics and processes. This paper presents a comparative review of AI algorithms that predict the readiness of by-product containers at slaughterhouses. The prediction allows the logistic planner to schedule the logistic resources earlier than usual. Consequently, the generated delay in the logistics can be reduced, or even eliminated. The trained models used real collected data from a processing facility for 10 months. Among several AI algorithms, both Decision Tree and Extra Trees regressors provided the lowest error. Then, the voting regressor of these two models provided better results and higher stability.

Keywords

Artificial Intelligence, Transit Time Prediction, Data Analytics, Feature Engineering

1. Introduction

Organic by-products from animal food production are usually not fit for human consumption. Nonetheless, they are used as raw materials to produce a wide range of commodities such as animal food, fertilizers, and biofuels, which in return, increases the sustainability of the entire food chain and improves environmental impact [1].

According to EU legislation EU 1069/2009 [2] and EU 142/2011 [3], the quality and category of the animal by-products depends on two main factors: the contents of the by-products and the age of the by-product. These two factors affect the types and quality of the produced commodities [4]. Therefore, it is necessary to optimize the logistics activities to maximize the quality of the by-products. One of the main challenges in this optimization problem is the narrow time window for the material before it starts decomposing. This time-critical nature increases the constraints, which in return, reduces the margin around the optimal solution.

With a substantial need for finding solutions to improve the environmental impact, the EU Commission is funding several research projects. One of these projects is titled Optimizing Production and Logistic Resources in the Time-critical Bio Production Industries in Europe (CLARUS) [5]. CLARUS project, funded by the EU Commission, intends to develop AI solutions for improving and sustaining the food industry. To validate the project goals, Honkajoki Oy – the leading animal by-product processing company in Finland- has been chosen in a use case involving logistics optimization. One scenario of this use case involves optimizing the selection of time-critical containers with the highest quality (i.e., category three material [3]) of animal by-products for transportation from slaughterhouses to Honkajoki's processing facilities. In this regard, this paper

Proceedings Acronym: : I-ESA 2024 12th International Conference on Interoperability for Enterprise Systems and Applications, April 10–12th, 2024, Crete, Greece

tuan.vu@tuni.fi (T. Vu); ali.jedariheidarzadeh@tuni.fi (A.J. Heidarzadeh); seyedamir.ahmadi@tuni.fi (S. Ahmadi); wael.mohammed@tuni.fi (W. M. Mohammed); mika.tuomola@honkajokioy.fi (M. Tuomola); jose.martinezlastra@tuni.fi (J.L. Lastra)

0000-0001-8615-0736 (T. Vu); 0009-0001-8943-6980 (A.J. Heidarzadeh); 0000-0002-3191-0504 (S. Ahmadi); 0000-0002-8364-3348 (W. M. Mohammed); 0009-0005-5869-2948 (M. Tuomola); 0000-0001-6227-3408 (J.L. Lastra)

presents the performance results of several machine-learning models trained on historical logistics data. The main objective of this research is to provide an empirical comparison of AI algorithms that provide accurate predictions of the availability of the material. Such a prediction may help the end users to react early, which in return, enlarges the narrow time window.

This paper is structured into several sections. The introduction section provides the context of the paper and the CLARUS project. Section 2 contains insights into by-product logistics optimization and the goal of forecasting the transit time. The approach developed to tackle the issue is explained in Section 3, while the preliminary results are presented in Section 4. Lastly, Section 5 describes the concluding remarks and potential next research steps.

2. Review of animal by-product logistics

Logistics is an integral part of supply chains and directly influences expenses [6]. Not only does optimized logistics improve supply chains' efficiency and enhance customer satisfaction, but it also leads companies toward greenness and sustainability. Better logistics means better resource allocation and less energy consumption and pollution [7]. Moreover, due to their deteriorating nature, food products differ from other types of material; hence, they require specific needs in their transportation [8].

In the case of Honkajoki Oy, logistics is of great significance since the material that is processed is highly time-critical. Raw material degrades gradually; hence, it should arrive at the factory for processing as soon as possible. Otherwise, the material quality would decrease to lower categories that require much more energy to process or be discarded due to the biochemical and chemical deterioration of the contents, especially in the case of category three animal by-products [3]. In the use case mentioned in this paper, category three chicken by-products are transported from three slaughterhouses to the Honkajoki processing facility by fleets of trucks. In a scenario, there can be multiple filled containers waiting for transport. To this end, an optimization algorithm is designed to analyze and guide operators in container selection at the slaughterhouse to maximize the quality and the number of category three containers. This optimization algorithm makes use of a machine learning model that predicts the container transit time from slaughterhouses to the Honkajoki factory yard instead of using average values.

3. Approach

3.1. Data collection

In the Honkajoki use case, data collection and management are arranged according to the system map as shown in Figure 1. Honkajoki has collected several years' worth of logistics and processing data and stored them on an Amazon AWS server (called Honkajoki Cloud). Historical logistics data used in the scenario described in this paper are collected from the Honkajoki Electronic Logistic System (HELOS) hosted within the Honkajoki cloud. The logistics data include container and truck data and timestamps of all logistics actions.

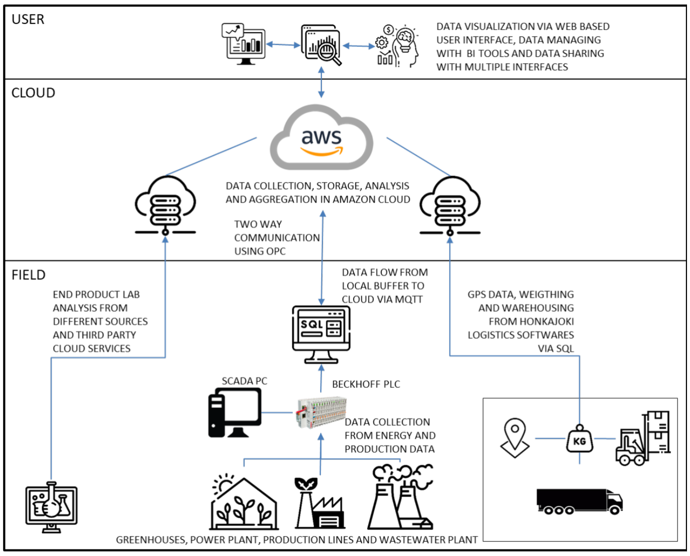

Figure 1: Honkajoki system map.

Figure 2 shows an example of a logistic dataset provided by Honkajoki.

containerID	slaughterhouse	Category	Weight	Filling start	Filling end	Pickup	Plate	Arrival
617	sh1	Sekatuote siipikarja luokka 3	13760kg	2020-12-30T15:48:00Z	2020-12-30T17:38:00Z	2020-12-31T08:41:00Z	truck1	2020-12-31T10:59:00Z
635	sh1	Sekatuote siipikarja luokka 3	12096kg	2020-12-30T07:02:00Z	2020-12-30T11:13:00Z	2020-12-31T08:42:00Z	truck1	2020-12-31T11:15:00Z
653	sh2	Sekatuote siipikarja luokka 3	13220kg	2020-12-31T06:10:00Z	2020-12-31T10:34:00Z	2020-12-31T12:23:00Z	truck2	2020-12-31T14:42:00Z
651	sh2	Sekatuote siipikarja luokka 3	16540kg	2020-12-31T05:49:00Z	2020-12-31T10:34:00Z	2020-12-31T12:24:00Z	truck2	2020-12-31T14:55:00Z

Figure 2: Historical logistics data example

3.2. Data modeling

The dataset used in this research holds information on containers, such as their raw material type, weight, and the slaughterhouse they have filled. Also, there are logistics-related attributes, e.g., the truck plate, the timestamp when a container finishes filling, and the timestamp when a truck reaches the Honkajoki yard.

Intuitively, the transit time of containers from slaughterhouses to the yard should depend on the time trucks leave slaughterhouses, which slaughterhouse they depart from, and the plate numbers identifying trucks. Weights of the containers are not considered since weighing containers takes place after reaching the Honkajoki yard. Consequently, the corresponding features in the preprocessing stage were selected and then encoded the string attributes, i.e., slaughterhouse name and plate number, to numerical values. Afterward, the timestamps were divided into four subparts, namely week number, weekday number, hour, and minute. Figure 3 shows five random rows of the dataset for training transit time predictors.

	SH	Plate	Leaving Week No.	Leaving Day	Leaving Hour	Leaving Minute
0	0	0	24	4	1	17
1	0	2	18	1	12	40
2	1	5	22	3	7	57
3	1	8	52	1	20	54
4	0	2	1	3	1	18

Figure 3: A brief overview of the dataset after preprocessing

The time difference in seconds between trucks leaving the slaughterhouse and arrival at the yard was used as labels to train the models. A sample of the transit time of containers from slaughterhouse 1

(SH1) recorded in the historical dataset from 5/12/2022 to 25/12/2022 and their departure timestamps are plotted in Figure 4.

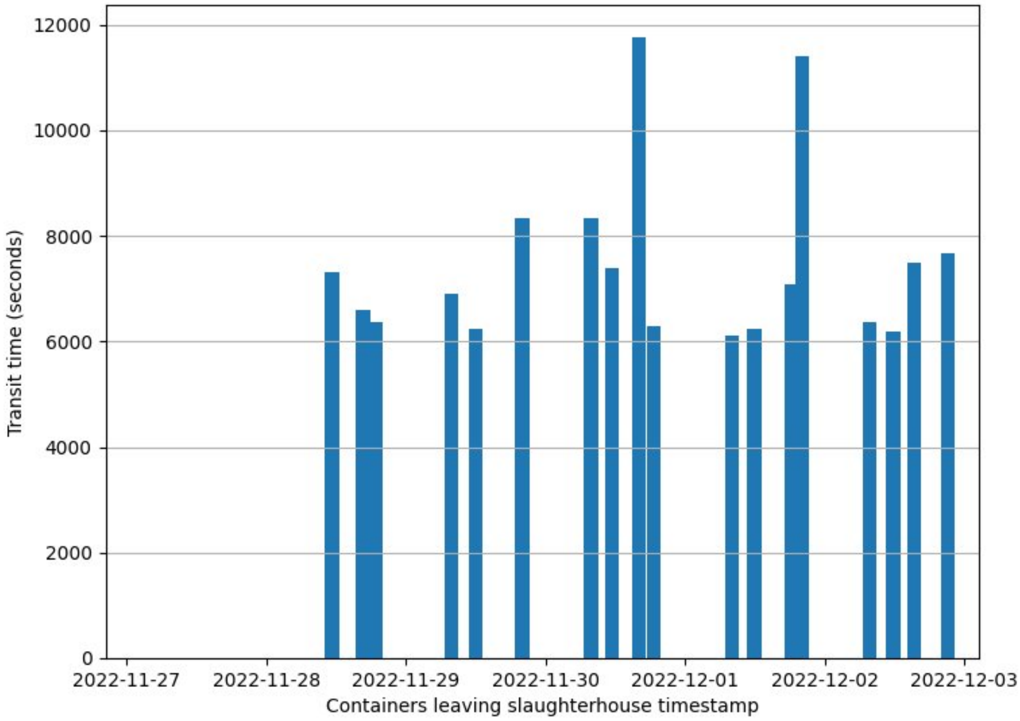

Figure 4: Container transit time plot.

3.3. Prediction model development

Predictive analytics uses historical data and statistics to analyze trends and predict or forecast the data. Predictive analytics is done by utilizing statistical algorithms and machine learning algorithms, allowing organizations to be proactive in situations in the future based on examining predicted data. Hence, predictive analytics has grown significantly, and multiple machine learning algorithms for various prediction tasks, including time-series forecasting, have been developed to improve the overall accuracy of the forecasted data [9] [10].

Commonly used models in time-series forecasting and predictive analysis, such as deep learning regression models, e.g., multi-layer perceptron (MLP) and long short-term memory (LSTM) neural networks, and ensemble learning algorithms, such as Random Forest regressor were considered for this use case. These models considered for testing are presented in Table 1.

Table 1
Machine learning algorithms used for testing.

Machine learning technique	Algorithm
Ensemble learning	Random Forest regressor
	Decision Tree regressor
	Gradient Boosting regressor
	Extreme Gradient Boosting regressor
	Support Vector Machine regressor
	Extra Trees regressor
	Voting regressor
Deep learning	LSTM neural network regressor
	MLP neural network regressor

4. Preliminary results

In this paper, ensemble learning models are created using existing algorithms from the scikit-learn Python library [13], except the Extreme Gradient Boosting regressor, while deep learning models use components from the TensorFlow library. Several models from the algorithms are created with multiple parameter configurations, e.g., different numbers of layers and neurons in the case of neural networks. The configurations yielding the best results based on Mean Absolute Error are shown in Table 2. Additionally, a linear regression model was used as the baseline for comparison.

The logistics data is randomly split into training and test sets with a ratio of 4:1, and all the models are trained and tested with the same dataset. The performance results of all the models are shown in Table 2.

Table 2: Machine learning algorithms used for testing.

Algorithm	Additional parameter configuration note	Mean Absolute Error (seconds)
Random Forest regressor	no maximum tree depth	491
Decision Tree regressor		381
Gradient Boosting regressor		753
Support Vector Machine regressor		832
Extreme Gradient Boosting regressor		464
Extra Trees regressor		285
Voting regressor	Extra Trees regressor, Decision Tree regressor	357
LSTM neural network model 1	51265 parameters	900
LSTM neural network model 2	200833 parameters	895
LSTM neural network model 3	84289 parameters	964
MLP neural network model 1	17550 parameters	985
MLP neural network model 2	67854 parameters	925
MLP neural network model 3	34062 parameters	895
Linear regression (baseline)	N/A	951

Table 2 shows that ensemble learning models perform better than deep learning models and the baseline model using the current dataset and input modeling method. The voting regressor combines Decision Tree and Extra Trees regressors, which already have low error margins, to produce final predictions with the lowest Mean Absolute Error. According to Table 2, the extra trees regressor performs the best; however, since there is randomness involved with this method, the voting regressor that combines the two best-performing models was chosen to ensure the stability of the predictions. Additionally, from testing results, while increasing the number of trainable parameters to the deep learning models can improve the forecasting results in some cases, the improvements do not scale well with the training cost of the models.

5. Conclusion

Processing animal by-products is a time-critical operation that requires minimizing any wasted time. The process itself can be well-planned. However, the transportation of the material from the slaughterhouses to the processing facility may generate delays and unplanned changes. Thus, predicting such disruptions in the logistic operations improves the overall result of the process of the by-product. As presented in this paper, AI- trained model on historical data can provide the needed prediction. As observed in this research, the voting regressor combining Decision Tree and Extra Trees regressors provides the lowest error with better performance in terms of stability. Future work may include better testing results from other prediction algorithms with different approaches to feature engineering. According to the requirements of the original scenario of the use case, the results from the optimization algorithm using data produced by prediction models will also be presented in the future.

Acknowledgements

This research has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No. 101070076. This number corresponds to the research project CLARUS which is titled as Optimizing Production And Logistic Resources In The Time-Critical Bio Production Industries In Europe.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] B. O. Alao, A. B. Falowo, A. Y. Chulayo and V. Muchenje, "The Potential of Animal By-Products in Food Systems: Production, Prospects and Challenges," *Sustainability,* vol. 9, no. 6, p. 1089, 2017.

[2] "Regulation (EC) No 1069/2009 of the European Parliament and of the Council of 21 October 2009 laying down health rules as regards animal by-products and derived products not intended for human consumption," 14 November 2009. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2009/1069/oj#. [Accessed 4 January 2024].

[3] "Commission Regulation (EU) No 142/2011 of 25 February 2011 implementing Regulation (EC) No 1069/2009 nd implementing Council Directive 97/78/EC," 26 February 2011. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2011/142/oj. [Accessed 4 January 2024].

[4] EFPRA, "Rendered products," [Online]. Available: https://efpra.eu/rendered-products/. [Accessed 2024 January 20].

[5] "CLARUS project," [Online]. Available: https://clarus-project.eu/. [Accessed 15 January 2024].

[6] A. Fredriksson and K. Liljestrand, "Capturing food logistics: a literature review and research agenda," *International Journal of Logistics Research and Applications,* vol. 18, no. 1, pp. 16-34, 2015.

[7] M. Christopher, Logistics and supply chain management, Pearson Uk, 2011.

[8] R. van Hoek, "Postponement and the reconfiguration challenge for food supply chains," *Supply Chain Management: an international journal,* vol. 4, no. 1, pp. 18-34, 1999.

[9] J. G. D. Gooijer and R. J. Hyndman, "25 years of time series forecasting," *International Journal of Forecasting,* vol. 22, no. 3, pp. 443-473, 2006.

[10] V. Kumar and M. L. Garg, "Predictive Analytics: A Review of Trends and Techniques," *International Journal of Computer Applications,* vol. 182, pp. 31-37, 2018.

[11] Mahboob Elahi, Samuel Olaiya Afolaranmi, Jose Luis Martinez Lastra and Jose Antonio Perez Garcia, "A comprehensive literature review of the applications of AI techniques," *Discover Artificial Intelligence,* vol. 3, 2023.

[12] M. S. N. Richard A. Davis, "Modeling of time series using random forests: theoretical developments," *Electronic Journal of Statistics,* vol. 14, pp. 3644-3671, 2020.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* vol. 12, pp. 2825--2830, 2011.