DECSUP-12791; No of Pages 12

Decision Support Systems xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates

Meditya Wasesa [a,b,c,*], Andries Stam [b], Eric van Heck [c]

[a] Indonesian Center for Logistics and Value Chains, Sarijadi Baru 15, 40152, Bandung, Indonesia
[b] Almende BV, Stationsplein 45, 3013 AK, Rotterdam, The Netherlands
[c] Rotterdam School of Management, Erasmus University, Burgemeester Oudlaan 50, 3062 PA, Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 23 August 2015
Received in revised form 10 October 2016
Accepted 29 November 2016
Available online xxxx

Keywords:
Predictive analytics
Gradient boosting model
Trajectory data
Vehicle telematics system
Seaport appointment system
Drayage operation

ABSTRACT

For drayage operators the service rate of seaports is crucial for organizing their container pick-up/delivery operations. This study presents a seaport service rate prediction system that could help drayage operators to improve their predictions of the duration of the pick-up/delivery operations at a seaport by using the subordinate trucks' trajectory data. The system is constructed based on three components namely, trajectory reconstruction, geo-fencing analysis, and gradient boosting modelling. Using predictive analytic techniques, the prediction system is trained and validated using more than 15 million data records from over 200 trucks over a period of 19 months. The gradient boosting model-based solution provides better predictions compared with the linear model benchmark solution. Conclusions and implications are formulated.

1. Introduction

In many parts of the world, road congestion near seaports is a common issue [1,2]. This very issue has led to a number of problems, namely the challenge of balancing resource utilization for the seaport operators, unproductive waiting time for the drayage trucks at over-utilized ports, queues of trucks increasing road congestion and generating excessive pollution, and so on [3]. The problems are becoming more severe as more shipping lines use bigger vessels [4,5] and as roads around port perimeters become more congested by commuter and freight traffic [6].

Infrastructure expansion initiatives are expensive, thus many alternatives have been proposed to mitigate the effects of congestion near seaports [2,7]. Existing road-traffic mitigation initiatives can be classified into two categories: diversion and non-diversion. Diversion initiatives, namely extended gate [8] and dry port initiatives [9,10], aim to divert the road commodity flow onto alternative channels such as rail or inland waterways. Not without consequences, diversion initiatives are associated with considerable implementation challenges such as business feasibility, public-private support, and infrastructure preparedness [10].

Non-diversion initiatives focus on improving the working condition of the seaport itself. In this category, popular initiatives include the extension of the seaport gate's opening hours and the improvement of the seaport gate appointment system [7]. The initiative to extend the seaport gate's opening hours aims to de-concentrate the peak load by offering more off-peak working hours. However, it is not easy to provide clear incentives to the drayage operators to use the new opening-hours alternative and to persuade both the consignors and the consignees to accommodate the extended schedule [7].

Alternatively, a gate appointment system can be used to monitor the arrival of trucks, mitigate load during a seaport's peak period, reduce road congestion, and improve resource utilization [7,11]. This initiative is less expensive than the gate extension initiative in terms of the human capital and land acquisition spending [7,12]. However, some articles [12, 13] have reported that existing systems deliver decent impact in mitigating the congestion of seaport roads. The systems were mainly used to retrieve information concerning commodities clearance status and were perceived as having minimum impact on improving container pick-up/delivery operations [11,14]. The negative issues with appointment systems [11,14] have resulted in low participation from drayage operators.

In this study, we focus on how drayage operators can apply predictive analytic techniques to their data assets to extract better insights and improve their operational decision making [15]. Doing this, we can circumvent the need to modify the design of the existing appointment system. We approach the problem from the drayage operators' perspective that seeks to minimize the loading/unloading time at seaports, and we consider the drayage operators' wealth of trajectory data, mined from the subordinate trucks' telematics system [16,17], as a valuable resource for evaluating a seaport's service rate. The objective

* Corresponding author.
E-mail addresses: m.wasesa@iclov.org (M. Wasesa), andries@almende.org (A. Stam), evanheck@rsm.nl (E. van Heck).

http://dx.doi.org/10.1016/j.dss.2016.11.008
0167-9236/© 2016 Published by Elsevier B.V.

of this study is to present a seaport service rate prediction system that uses the trajectory data communicated through the drayage trucks' telematics devices.

In Section 2, we review the literature on predictive analytics and seaport appointment systems including the potential usage of predictive analytics. Section 3 presents the service rate prediction system framework and its step-wise approach. In Section 4, we present the background of the case study in the Port of Rotterdam and the used datasets and Section 5 presents and evaluates the results of the service-rate prediction system. Section 6 concludes and presents limitations and future research.

2. Literature review

The use of predictive analytics [18] to improve business operations has received increasing attention from both the research and business communities [19,20]. Technological artefacts that can access and synthesize large volumes of data to produce useful operational insights are in demand [19,20]. On the technological side, the ubiquity of powerful sensing technologies [21] and the increasing use of predictive analytics [18] have stimulated the development of novel and better predictive systems. Despite the increasing development of predictive systems, studies that focus on the container pick-up/delivery operations context are hard to find.

2.1. Related literatures on predictive analytics

Predictive analytics strengthen business intelligence (BI) systems' feature in providing quality inputs to support operational decision-making process by delivering the right information at the right moment, at the right place, and in the right forms [19,20,22]. Note that, the BI systems concept can be seen as similar to the integrated decision support systems concept which positions data warehouses as inseparable component of intelligent decision support systems [23]. Predictive analytics enhance the data retrieval capabilities of BI systems through statistical models or empirical methods that are aimed at creating and assessing empirical predictions [15,18]. Unlike the conventional statistical modelling approach, predictive analytics aim to develop a prediction inference with pragmatic business relevance rather than mere data description/visualization for exploration or to test theories [15,18].

More recently, the application of predictive analytics to novel data sources (from sensor to social media data) has received increasing interest [19,20]. While predictive analytics development using data contents from database management systems (DBMS) or the web are plentiful, a study indicated that the underlying mobile analytics and location, and context-aware techniques for collecting, processing, analyzing, and visualizing these mobile and sensor data are largely unexplored [20]. Making use of the vehicle telematics trajectory data, this study responds to the high demand for predictive analytics that process sensor-based data. While the development of data mining techniques to analyzing vehicles trajectory datasets has been well researched [24–26], building predictive analytics to support operational decision making is still an under-researched field [27–29].

2.2. Seaport appointment systems and predictive analytics

Many stakeholders perceived seaport appointment systems to yield a minimal improvement to container pick-up/delivery operations [12,13]. Both seaports and drayage trucking companies are two primary users of the appointment system. Port authorities, local governments, and the people living around seaports are some of the secondary stakeholders who receive negative externalities from the system's deficiencies. Dissatisfaction with appointment system design has evoked many further research initiatives that aim to improve upon the system's design limitations [1,12,30–32]. Since the design of the appointment system concerns the interest of multiple stakeholders, different studies have analyzed the design factors most relevant to the interests of specific stakeholders [30,31,33–35].

Some studies analyzed the design from the seaport's point of view [30,36], focusing on the impact of limiting truck arrivals at seaports on truck turn time and yard crane utilization [37], the impact of adjusting the number of trucks allowed to enter the seaport per period [30], the number of truck appointments to offer and resource allocation [38], the seaport's service line configuration [39], the truck arrival disruption management concern [33], and so on. Some studies analyzed the design from the drayage operators' perspective [32], focusing on how access capacity and the parameters of appointment time windows influence the productivity of drayage trucks [31], the definition of optimum pick-up/delivery time window parameters [40], and so on [35]. Other studies considered information visibility, specifically the fact that seaports have limited information about trucks before they arrive and vice versa [32,41]. It is not trivial to achieve clear benefit propositions for all stakeholders (particularly seaports and drayage trucking companies) when modifying the design of the appointment system. Despite extensive studies on appointment system design, the affected stakeholders do not always agree to apply the recommended changes. Modifying an existing appointment system may require radical adjustments to stakeholders' existing operations. Factors such as the stakeholders' view of the prospective costs and benefits, strategic considerations, aversion to risk, and unwillingness to invest can prevent them from supporting proposed alterations to the appointment system [6].

To circumvent the need to modify the appointment system's design, companies can explore the opportunity to use their internal data assets to extract better insights and improve their operational decision making using predictive analytics [15,19,20]. To the best of our knowledge, this approach has not been applied in any research initiative aiming to improve the container pick-up/delivery operations, especially the ones that correlated with the appointment system initiatives. As shown in Table 1, previous studies focused on predicting the productivity performance of the seaports seaside operations namely, container-handling

Table 1
Related articles on seaport productivity predictions.

Article	Predicted variables	Predictors	Model
[42]	Cargo throughput (TEUs/year); ship working rate (TEUs/hour)	Number of cranes; number of berths; number of tugs; service delay (hour; port area (m^2)); labor (employees)	Data envelopment analysis
[43]	Cargo throughput (TEUs/year)	Cargo throughput (TEUs/year); foreign trade value (million USD); port tariff (USD)	Structural vector error correction model
[44]	Seaport efficiency scale (0–1)	Port length (m); terminal area (ha); number of quayside gantry; number of yard gantry; number of straddle carrier	Data envelopment & stochastic frontier analyses
[45]	Cargo throughput (TEUs/month)	Cargo throughput (TEUs/month)	Genetic programming; decomposition approach; seasonal auto regression integrated moving average
[46]	Cargo throughput (TEUs/year)	Gross domestic product (CNY); fixed assets investment (CNY); imports & exports value (USD); industrial output (CNY); primary, secondary, and tertiary industrial value (CNY); population (people); total consumer goods retail sales (CNY); total freight, highway freight, and railway volume (tons)	Robust V-support vector regression; simulated annealing particle swarm optimization; Multivariable adaptive regression splines

Please cite this article as: M. Wasesa, et al., The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates, Decision Support Systems (2016), http://dx.doi.org/10.1016/j.dss.2016.11.008

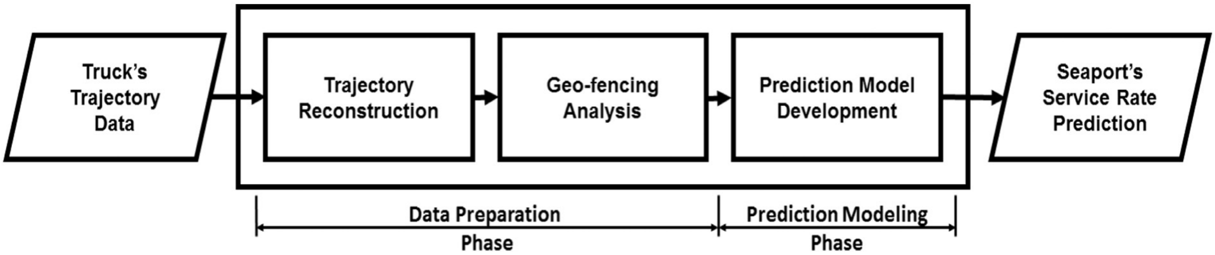

Fig. 1. The framework of the service rate prediction system.

throughput or ship working rate values [42–46]. In general, those studies used yearly/monthly statistical archives to build predictions to supporting strategic or socioeconomic decisions. In contrast, this study focuses on the seaports' landside productivity using sensor-based trajectory dataset that is updated every few minutes to supporting drayage operators' operational decisions.

3. The service rate prediction system

Fig. 1 presents the overview of our service rate prediction system which provides a step-wise overview on how it could help the drayage operators to make use of their trucks' trajectory data asset for predicting a seaport's service rate. As portrayed, users only need to provide the truck's trajectory data inputs and the prediction system will produce the seaport's service rate outputs in return. Note that the proposed framework can also be applied to predict the service rate of any service station other than seaports.

This study departs from the role and limitation of the existing seaport appointment system in facilitating the container pick-up/delivery operations. In general, a seaport appointment system facilitates the pre-notification procedure, a prerequisite for drayage trucks to execute the container pick-up/delivery operation. The pre-notification procedure aims to prevent unsynchronized pick-up/delivery operations in which trucks wait at the seaport for long periods due to many possible issues such as container absence and documents clearance problems.

The existing pre-notification formalities are in accordance with the United Nations global rules for Electronic Data Interchange for Administration, Commerce and Transport/UN-EDIFACT [47]. The drayage operator initiates the pre-notification process by sending the pick-up/delivery permission request, known as the COPINO message, to the seaport. The seaport will then verify the information submitted by the drayage operator, the container's presence, the customs clearance status, and so on. Subsequently, a reply (APERAK) message will be sent to the drayage operator indicating the COPINO message approval or rejection. The drayage operator's truck can conduct the pick-up/delivery at the predefined port and date only after receiving an approval of the COPINO message. If the COPINO message is rejected, the operator will review the reasons, carry out corrective actions, and re-submit a new COPINO message.

The period between the sending of the COPINO message and the receipt of the APERAK approval is the pre-arrival phase, while the period when the drivers travel to the seaport is the travel phase, and the period when the container loading/unloading execution happens is the on-arrival phase (see Fig. 2). To speed up order delivery to the consignee/consignor, the drayage operators often dispatch their trucks to the seaport even before receiving the APERAK approval so that the pick-up/delivery execution can be conducted directly when the approval is received.

The APERAK approval only contains information about a container's availability, namely whether a container can be picked up/delivered on a specific date. This information is not enough to assess the seaport's service rate, specifically the duration of the pick-up/delivery operation [14]. Nevertheless, service rate information is important to conduct better vehicle routing that will increase trucks' productivity [48–50]. Since drayage operators are commissioned based on the number of finalized pick-up/delivery services and large shares of time-dependent operational costs (such as the truck driver salary, administrative expenses) are allocated for the operation, finalizing as many orders as quickly as possible is important. Spending excessive time either in the travel phase or in the on arrival phase has to be avoided [49,50].

To provide the drayage operators with the required seaport's service rate predictions, we propose a service rate prediction system. We only use the truck's trajectory data as the system's primary input. Prior to

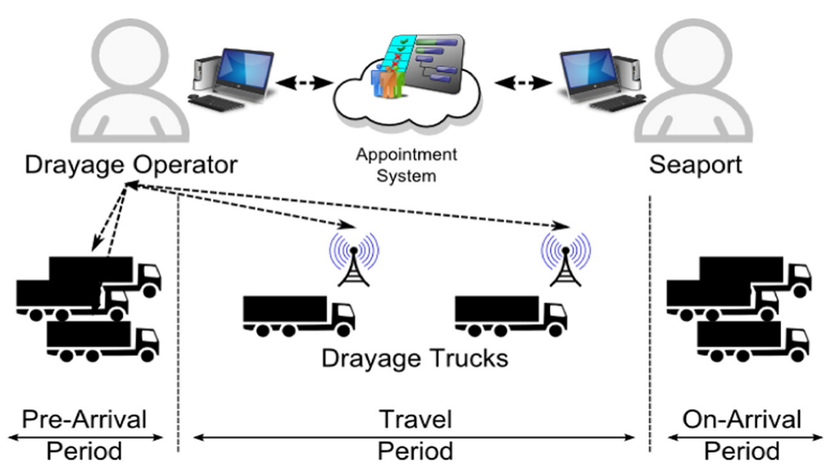

Fig. 2. The container pick-up/delivery operation.

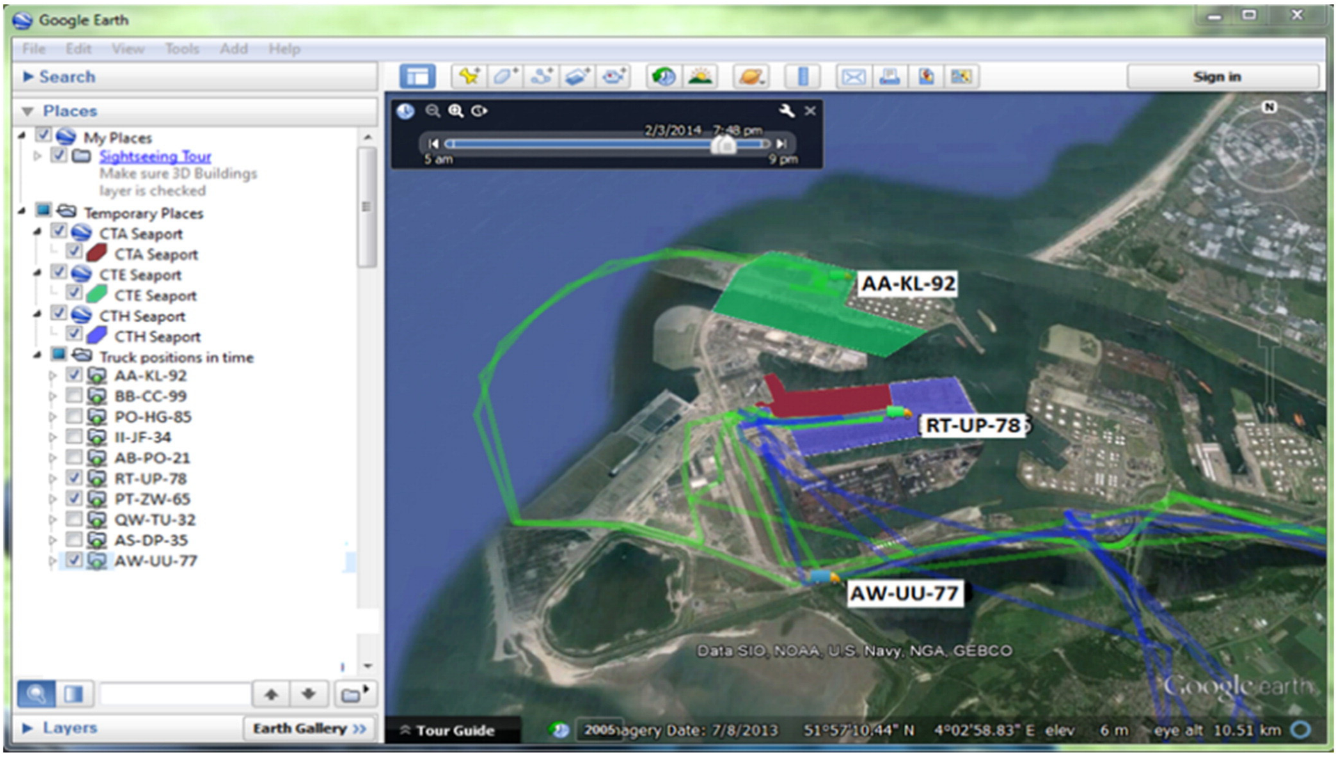

Fig. 3. The outcome of trajectory reconstruction and geo-fencing analyses.

the development of the prediction model, the trajectory data undergo a preparation process which includes trajectory reconstruction and geo-fencing analyses.

3.1. Trajectory reconstruction

Drayage operators normally apply the temporal sampling strategy to monitor subordinate trucks [27], such that each batch of the telematics data is uploaded from the truck's telematics system at a pre-defined time interval. The data logs can then be used to reconstruct the trucks' position [27,51]. The higher the sampling frequency (or the shorter the message sending intervals), the more accurate the results of the trajectory reconstruction will be. However, higher sampling frequency is not always possible owing to high communication load and high energy consumption [27].

As shown in Fig. 3, the output of the trajectory reconstruction process plots the trucks' historical position on a map. Note that the resolution of the reconstruction outcome is of higher quality as we magnify the trajectory reconstruction result. The trajectory reconstruction process is essential to understanding the movement of each truck. With this process, it is easier to get an overview of a drayage operator's service coverage, to analyze detailed mobility data of each truck, and to target some locations with the potential for further analysis.

3.2. Geo-fencing analysis

To assess a seaport's service rate performance, we apply the geo-fencing technique adopted from the wireless network research field [52]. Originally, this technique was used to define the service areas of Wi-Fi access points to a specified region [52]. In this study, we apply the geo-fencing technique and target the vehicle trajectory GPS signal area of the reviewed port (see Fig. 3). Next, we measure the duration the truck remained in the reviewed area, specifically we record the truck's time of arrival and time of departure at the analyzed seaport. Table 2 shows the expected output format of this analysis.

Subsequent to the geo-fencing analysis, we can calculate the seaport's service rate. The service rate value is inferred from the time the truck remained at the seaport area. The service rate performance $\varphi_{t_x}{}^i$ of a certain seaport i at a specific time period t as measured by the respondent truck x, can be computed by subtracting the time of departure $t_{departure}$ and the time of arrival $t_{arrival}$ of the respondent truck x. The t notion in the $\varphi_{t_x}{}^i$ is set equal to $t_{arrival}$. The seaport's service rate $\varphi_{t_x}{}^i$ is then defined as follows:

$$\varphi^i_{t_x} = \theta^i_{t_{departure_x}} - \theta^i_{t_{arrival_x}} \tag{1}$$

Note that we characterize the service rate φ based on three identifiers: the seaport's identity i; the time at which the service execution begins t; and the respondent truck's identity x. The first index i, the seaport's identity, is used to differentiate the performance of one seaport from another. The second index t is included since we conjecture that the service rate magnitude will vary based on the timing of service execution. The inclusion of the time index t emphasized the importance of the timing of the service execution aspect when constructing the vehicle routing plan. The service rate at one seaport may be better than another at a specific time, but the same seaport may not provide a competitive service rate at other times.

The last index x refers to the identity of the respondent truck whose telematics data were used to measure the seaport's service rate. It is often the case that more than one truck visited the seaport i at the same time period t. However, the measurement results may not be the same. Denoting $X = \{x_1, x_2, \dots x_n\}$ as the valid respondent trucks

Table 2
The format of the geo-fencing analysis output.

Company	Truck ID	Time of arrival		Time of departure	
		Date	Time	Date	Time
LDH	AA-KL-92	10/10/2016	13:25:36	10/10/2016	13:45:22
SUP	AS-WR-1	10/10/2016	13:39:41	10/10/2016	14:09:17

Please cite this article as: M. Wasesa, et al., The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates, Decision Support Systems (2016), http://dx.doi.org/10.1016/j.dss.2016.11.008

that visited the seaport i at a specific time period t, we infer the seaport's service rate φ_t^i by calculating the arithmetic mean from all measurement records of the respondent trucks. At the initial phase (i.e. $t = 0$), $\varphi_{t=0}^i$ will be set to a constant value of c^i that will be entered by the user according to the user's estimation of seaport i service rate value. In a few cases, it could be that the operator's data logs do not contain any record of a truck respondent that visited the seaport i at a specific time period t (i.e. $X = \emptyset$). In this case, we infer the service rate value at time t from the service rate value of the preceding time period to capture the inertia effect from the seaport's service momentum at time *t-1*. As an example, the service rate of a seaport at time $t = 13$ ($i.e. \varphi_{t=13}^i$) will be extrapolated from the service rate of the same seaport at the preceding time $t = 12$ ($i.e. \varphi_{t=12}^i$) so that $\varphi_{t=13}^i = \varphi_{t=12}^i$. Formally, we state the seaport's service rate value as follows:

$$\varphi_t^i = \begin{cases} c^i & , \quad t = 0 \\ \varphi_{t-1}^i & , \quad X = \emptyset \\ \frac{1}{n}\sum_{x=1}^{n}\left[\theta^i_{t_{departure_x}} - \theta^i_{t_{arrival_x}}\right] & , \quad |\boldsymbol{X}| \geq 1 \end{cases} \tag{2}$$

3.3. Prediction model development

To predict a seaport's service rate, we adopt the generalized additive model that falls under the regression framework category. The generalized additive model technique [53] is preferable to generalized linear models because it allows us to make inferences about associations between predicted variables and predictors without including any parametric restrictions on the associations [54]. We define the formalization of the seaport service rate predicted value as follows:

$$\overline{\varphi_t^i} = \alpha_i(t) + \beta_i\left(\gamma_t^i\right) + \varepsilon_{it} \tag{3}$$

where:

- $\overline{\varphi_t^i}$ denotes the predicted value of the service rate at the seaport i at time t;
- $\alpha_i(t)$ models the temporal effects predictor (the monthly, daily, hourly, and quarterly effects);
- $\beta_i(\gamma_t^i)$ models the inertia effects predictor of the seaport's recent performance; The γ_t^i vector consists of recent records of the service rate performance $\varphi_{t-z}^{i^{\cdot}}$, the number of arriving trucks $\zeta_{t-z}^{i^{\cdot}}$, and the number of departing trucks $\delta_{t-z}^{i^{\cdot}}$; and
- ε_{it} refers to the prediction model's error.

Discussing the first predictor, we add the temporal effects $\alpha_i(t)$ since a seaport's service rate varies over time in the real world and we conjecture that the reversal of a seaport's performance can drive the reversal of the drayage operator's seaport preference. We translate this time-based effect into four attributes, namely the monthly, daily, hourly, and quarterly effects. The monthly effect was modelled with factor variables, adopting the month name (January, February, etc.). The daily and the hourly effects were also modelled with factor variables. The coding for the daily effects adopts the standard day values (Monday to Sunday) and the coding for the hourly effect follows the natural 24 h discretization of a normal day. The last attribute, the quarterly effect, is coded as a factor variable and computed by discretizing the 24 h period in 15 min increments. For example any service execution that started between 00.00 and 00.15 will have a quarterly index value of $t = 1$, any service execution that was started between 00.15 and 00.30 will have a quarterly index value of $t = 2$, and so on where $t \in T$ and $T = \{1, 2, \ldots, 96\}$.

By introducing the inertia effects $\beta_i(\gamma_t^i)$, we incorporate the momentum of the seaport's recent performance. Based on our previous experience in handling prediction tasks based on panel data, the inclusion of the inertia effects can significantly increase the prediction performance. The travel behavior's inertia effect has been also recognized as an important factor for modelling transport demand [55]. We translate the effects into three attributes, namely the historical trace of the seaport's service rate $\varphi_{t-z}^{i^{\cdot}} = \{\varphi_{t-z}^i, \ldots, \varphi_{t-2}^i, \varphi_{t-1}^i\}$, the historical trace of the seaport's arriving trucks $\zeta_{t-z}^{i^{\cdot}} = \{\zeta_{t-z-1}^i, \ldots, \zeta_{t-2}^i, \zeta_{t-1}^i\}$, and the historical trace of the seaport's departing trucks $\delta_{t-z}^{i^{\cdot}} = \{\delta_{t-z-1}{}^i, \ldots, \delta_{t-2}{}^i, \delta_{t-1}^i\}$. We incorporate the service rate trace factor φ_{t-z}^i to capture the seaport's recent performance in handling container pick-up/delivery requests. The prior numbers of arriving and departing trucks are incorporated as a mean for inferring the number of trucks inside the seaport area.

For the prediction task we opt for the gradient boosting method [56,57], a machine learning technique that constructs a regression prediction model by combining weak prediction models into an ensemble [15,58]. "Gradient boosting constructs additive regression models by sequentially fitting a simple parameterized function (base learner) to current "pseudo"-residuals by least squares at each iteration. The pseudo-residuals are the gradient of the loss functional being minimized, with respect to the model values at each training data point evaluated at the current step" [36,p. 367]. As a boosting algorithm, the model is chosen due to its strong prediction performance records [57,59] that can be associated with its robustness towards overfitting cases [60].

Applying the model to Eq. (3), we can re-write the seaport service rate prediction in the following form:

$$\overline{\varphi_t^i} = F_i(\psi_t) + \varepsilon_{it} \tag{4}$$

ψ_t consists of all potential variables within the temporal effects $\alpha_i(t)$ and the inertia effects $\beta_i(\gamma_t^i)$ that can be incorporated as predictors in the final gradient boosting model. By learning from the supplied training dataset that consists of actual ψ_t and φ_t^i values, the goal is to find the best function F_i: $R^d \to R$ that minimizes the prediction loss function. In this study we opt for the root mean squared error (RMSE) measure for the loss function [15,18]. In essence, the RMSE indicates the sample standard deviation of the differences between the actual service rate values and the values produced by the prediction model. Thus, the prediction function F_i based on the training set $\{(\varphi_t^i, \psi_t)\}_{t=1}^T$ can be written as follows:

$$\overline{\varphi_t^i} = \arg\min \frac{1}{T}\sum_{t=1}^{T}\left[\varphi_t^i - F_i(\psi_t)\right]^2 \tag{5}$$

The gradient boosting method approximates the prediction function $F_i(\psi_t)$ in a sequential manner. We denote $F_i^{(m)}(\psi_t)$ as the estimation of $F_i(\psi_t)$ at the m-th iteration, where $m = 0, 1, 2, \ldots, M$. The approximation starts with $F_i^{(0)}(\psi_t) = \overline{\varphi_t^i}$, where for the first iteration (i.e. $m = 0$) the value of $\overline{\varphi_t^i}$ uses the mean value of the service rate performance at the seaport i at time t. Subsequently, the model can be updated using:

$$F_i^{(m)}(\psi_t) = F_i^{(m-1)}(\psi_t) + \nu.\ h_m(\psi_t, Ш_m) \tag{6}$$

Whereas we denote $h_m(\psi_t, Ш_m)$ as the weak learner estimate at the m-th iteration with parameters $Ш_m$ and we denote $\nu \in [0, 1]$ as the shrinkage parameter. Given the approximation of $F_i^{(m-1)}(\psi_t)$, each additional term $h_m(\psi_t, Ш_m)$ can be computed by differentiating the loss function with the prediction $F_i(\psi_t)$ function:

$$\begin{aligned} u_t^m &= -\left[\frac{{}^1/{}_2 d\left[\varphi_t^i - F_i(\psi_t)\right]^2}{dF_i(\psi_t)}\right]_{F_i(\psi_t) = F_i^{(m-1)}(\psi_t)} \\ &= \left[\varphi_t^i - F_i{}^{(m-1)}(\psi_t)\right] \end{aligned} \tag{7}$$

Please cite this article as: M. Wasesa, et al., The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates, Decision Support Systems (2016), http://dx.doi.org/10.1016/j.dss.2016.11.008

Table 3
Specification of the truck telematics system's data.

Attribute	Variable		Data type
Record identification	No[a]		Integer
Timestamp	Timestamp[a]		Date and time
Truck identification	License Plate[a]		String
	Affiliated company[a]		String
	Driver's name		String
	Truck's capacity		Integer
Truck status	Location[a]	Longitude[a]	Double
		Latitude[a]	Double
	State	Loaded/empty	String
Destination	Location	Longitude	Double
		Latitude	Double
	Estimated time of arrival		Date and time

[a] Minimum data specification.

Eq. (7) produces the direction of the steepest descent step. Furthermore, a regression analysis is applied on $\{u_t^m, \psi_t\}_{t=1}^T$ by the weak learner as follows:

$$Ш_m = \arg\min \sum_{t=1}^{T} \left[u_t^m - h_m(\psi_t, Ш_m)\right]^2 \tag{8}$$

The $h_m(\psi_t, Ш_m)$ value is selected to estimate the prediction error of the prior model $F_i^{(m-1)}(\psi_t)$. Thus, the final solution can be written as follows:

$$F_i(\psi_t) = F_i^{(M)}(\psi_t) = h_0(\psi_t) + \sum_{m=1}^{M} v.h_m(\psi_t, Ш_m) \tag{9}$$

Note that the $F_i(\psi_t)$ function is continuously updated by the addition of $v.h_m$ component at the m-th iteration whereas the hyper-parameter M denotes the maximum number of adopted components, which will prevent overfitting cases.

4. Case study

With an annual throughput value of 465 million cargo tons, the Port of Rotterdam is currently the busiest seaport in Europe. Serving approximately 30,000 seagoing vessels and 110,000 inland vessels every year, the Port of Rotterdam is home to at least 12 container seaports and more than a hundred drayage operators [61]. As a case study, we analyzed the service rate of containers pick-up/delivery operations at three anonymous container seaports. The selected seaports are some of the most prominent seaports in the region in terms of containers throughput value. To build the prediction models, we use vehicle telematics data from three different drayage operators that visit the selected terminals regularly. Two respondent operators focus on providing transportation business only while the other one provides richer spectrum of services namely transportation, warehousing, global freight forwarding, etc. In this study, we analyze the operators' trucks that conduct pick-up/deliver services for European clients located in the Netherlands, France, Germany, Switzerland, and Spain.

4.1. Truck telematics system data

Noting the appointment system's limitation in terms of information content, the availability of a large amount of truck telematics system data [16] is an alternative to assess the seaport's service rate. The telematics system is used by drayage operators to monitor and communicate with their subordinate trucks. The board computer mounted on the truck's dashboard is a visible component of the system. In operationalizing the monitoring and the communication tasks, data are exchanged. The dataset contains many attributes such as the record identification number, the timestamp of the data recording, the truck's identification, the destination's location, and so on (see Table 3).

Each company may have a different policy about the information attributes that must be monitored. At a minimum, a drayage operator will monitor the trucks' position [17,62]. This trajectory information consists of the following: the record identification number; the data recording timestamp; the truck's license plate number; the truck's affiliated company; and the truck's location, specifically longitude and latitude data.

For this study, we imported >15 million real-life data records logged from >200 drayage trucks. The data were retrieved from the drayage operator's vehicle telematics database. The imported dataset contains trajectory records for a period of more than one and a half years, during which each truck transmitted an average of 130 messages per day. Within the dataset, a few trucks were found to have inconsistent data-updating behavior. To improve the data quality, we only considered data records from the trucks with frequent and consistent data updates. For the analysis, we used two filtered datasets with different message updating threshold rates of at least 15 min and 7.5 min (see Table 4). In our dataset only three out of 202 trucks transmitted data updates every 5 min, thus we cannot set the updating-rate threshold below 7.5 min.

The dataset containing records from trucks with an updating period of 15 min or less provided nearly 14 million lines of high quality data from 119 respondent trucks. On average, each truck transmitted nearly 200 messages per day, or one message every 7.3 min. The dataset with the 7.5 min updating threshold rate provided nearly 8 million data records from 54 respondent trucks. In the second dataset, on average, each truck transmitted nearly 250 messages per day, or one message every 5.9 min.

4.2. Analysis

To apply the proposed framework to the dataset, we first apply the geo-fencing analysis to the dataset by treating the analyzed seaport area as the bounding area. In general, all seaport bounding areas are polygon-shaped. For each seaport, one geo-fencing analysis output sheet will be produced. Subsequently, we transform each geo-fencing output sheet into the modelling sheet format (see Table 5). Since we evaluate six different prediction model variants, we prepare six modelling sheets from each geo-fencing output sheet. Each modelling sheet has a different set of target and predictors variable pairs $\{(\varphi_t^i, \psi_t)\}_{t=1}^T$. For model 0, we only include the temporal effect variables as the predictors

Table 4
Truck telematics system data description.

Variable	Metric	Raw data	Filtered data	
			Update rate	
			15 min	7.5 min
Number of records	Count	15,314,614	13,918,940	7,800,648
Data time span	Min	21/06/2012 08:29:04 GMT + 1	21/06/2012 08:29:05 GMT + 1	21/06/2012 08:29:36 GMT + 1
	Max	03/02/2014 15:59:44 GMT + 1	03/02/2014 15:53:24 GMT + 1	03/02/2014 15:53:24 GMT + 1
Days	Count	592	592	592
Number of trucks	Count	202	119	54
Daily message sent per truck	Average	128.07	197.58	244.01

Table 5
The format of the modelling sheet.

Construct	Attribute	Function
Temporal effects	Month	Predictor
	Day	Predictor
	Hour	Predictor
	Quarter	Predictor
Inertia effects	Previous service time (t-n)	Predictor
	Previous arriving trucks (t-n)	Predictor
	Previous departing trucks (t-n)	Predictor
Performance	Service time	Target

$\overline{\varphi_t^i} = \alpha_i(t) + \varepsilon_{it}$ where $\psi_t = (month_t, day_t, hour_t, quarter_t)$. For model 1, we add the inertia effect for the first time $\overline{\varphi_t^i} = \alpha_i(t) + \beta_i(\gamma_t^i) + \varepsilon_{it}$ where $\psi_t = (month_t, day_t, hour_t, quarter_t, \varphi_{t-1}, \zeta_{t-1}, \delta_{t-1})$. In model 2, we add the inertia effects to the second degree as predictors $\psi_t = (month_t, month_t, day_t, hour_t, quarter_t, \varphi_{t-1}, \zeta_{t-1}, \delta_{t-1}, \varphi_{t-2}, \zeta_{t-2}, \delta_{t-2},)$ and so on. In this study, the addition of the inertia effects goes until model 5, where we translate the predictors ψ_t as $(month_t, day_t, hour_t, quarter_t, \varphi_{t-1}, \zeta_{t-1}, \delta_{t-1}, \ldots, \varphi_{t-5}, \zeta_{t-5}, \delta_{t-5},)$.

In constructing the prediction models, we apply *k-fold* cross validation method [15,18] with $k = 10$. The method offers lower variance than the simplistic single hold-out cross validation method and offers faster computation time compared to the leave-one-out cross validation method especially when coping with a high volume dataset like ours [15]. Since we analyze six different target and predictor variable pairs, we construct and evaluate six gradient boosting prediction models (GBM_0: GBM_5) and six generalized linear models (LM_0: LM_5) for each seaport. Note that the generalized linear models [15] with the same target and predictors variable pairs are used as benchmarks for assessing the performance of the gradient boosting model.

5. Results and discussion

This section presents and evaluates the case study's result of the service-rate prediction system that is developed using the step-wise approach introduced in Section 3, namely the trajectory reconstruction, the geo-fencing analysis, and the prediction model development.

5.1. Trajectory reconstruction

In Fig. 4 we depict an outcome of the trajectory reconstruction process from a sample dataset (April – mid-June 2013). Through this analysis, one can plot the trucks' historical activity to better understand the trucks trajectory and service coverage during different periods (see Fig. 4c, d, and e). The heat map plot (see Fig. 4a) yields better insight

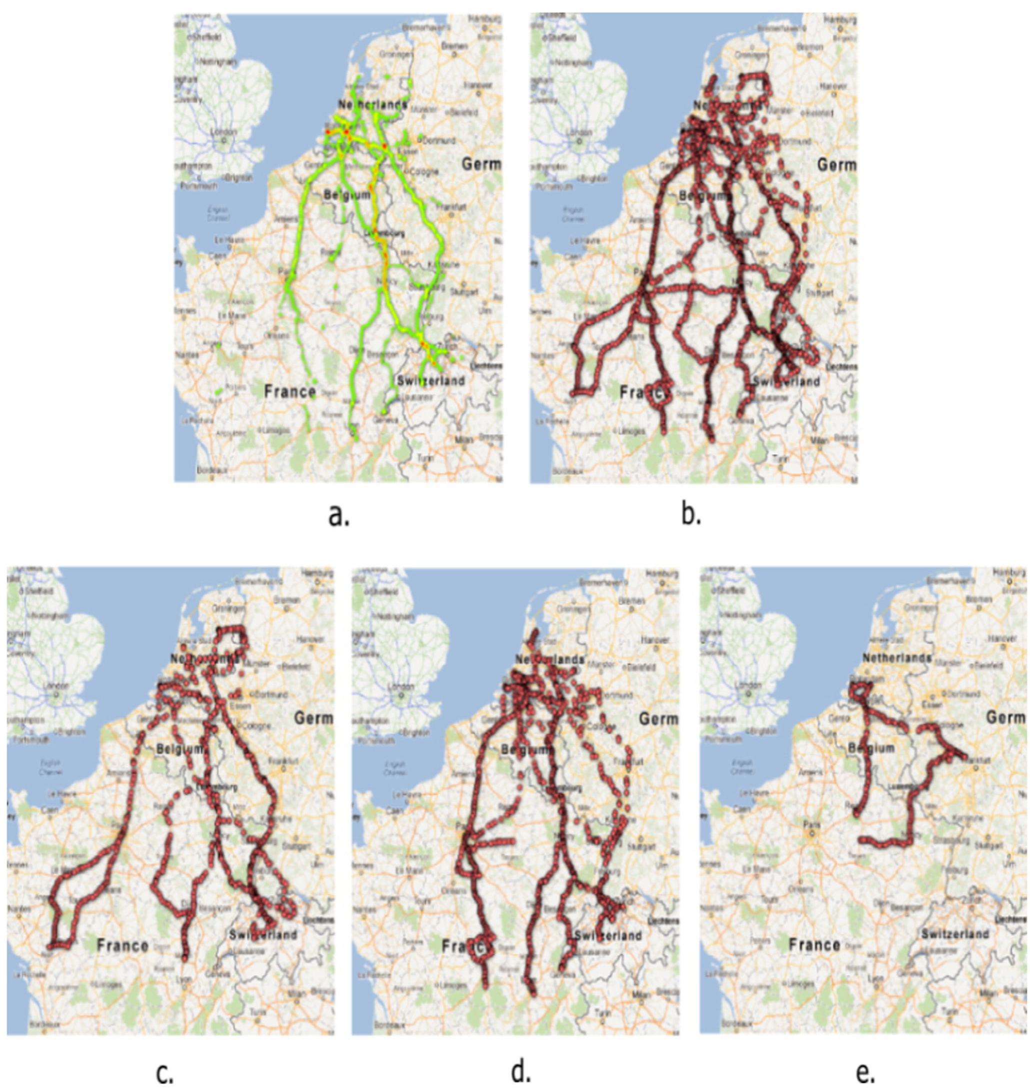

Fig. 4. The trajectory reconstruction outcome, (a) heat-plot (April–June), (b) dot-plot (April–June), (c) dot-plot (April), (d) dot-plot (May), (e) dot-plot (June).

into the mobility concentration of a drayage operator's subordinate trucks mainly around the Rotterdam area where the company's headquarters and its main costumers are located.

5.2. Geo-fencing analysis

Subsequent to the trajectory reconstruction process, we analyze the service rate performance of three seaports in the Rotterdam area, anonymized as CTA, CTE, and CTH. Each seaport is marked with different color namely, red (CTA), green (CTE), and blue (CTH). The application of the geo-fencing analysis to the trucks' trajectory data will produce one output sheet for each corresponding seaport. Depicting the results of the geo-fencing analysis in Fig. 5, we notice that each seaport has unique service characteristics and the seaport service rate profile varies over time. At different hours, days, and months of service execution, the seaports perform differently.

From Fig. 5 we can observe that the CTA generally performs better with a lower service time than the competitors. Looking at the hourly performance, all seaports tend to deliver longer services between 11:00 and 17:00 than any other period. Looking at the monthly performance, July is the time when all seaports perform the best service rate. We can see that at Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday the CTA performance is better than its competitors. However, the competitors perform better at Sunday (day 0). During this period, it is recommended to conduct pick-up/delivery service at seaports with a more competitive service rate. The findings are in line with our conjecture, namely that the service rate does indeed vary over time and the reversal of a seaport's performance can in fact drive the reversal of the preference for a particular seaport.

5.3. Prediction model development

In Table 6, we present the performance of the constructed prediction models in terms of the model's predictive power in RMSE form [15]. The RMSE indicates the sample standard deviation of the differences between the actual service rates and the predicted values (see Section 3.3. Eq. (5)).

For comparative purposes, we also included the descriptive statistics measures (mean and standard deviation) of each seaport's service rate. In line with the inferences that were made, we notice that in general the CTA is the best performing seaport, calculated based on the 15 min threshold data. Not only was the CTA found to have the shortest average service rate (CTA = 31.16 min compared to CTE = 39.36 min and CTH = 50.83 min), the CTA also has the lowest service rate deviation

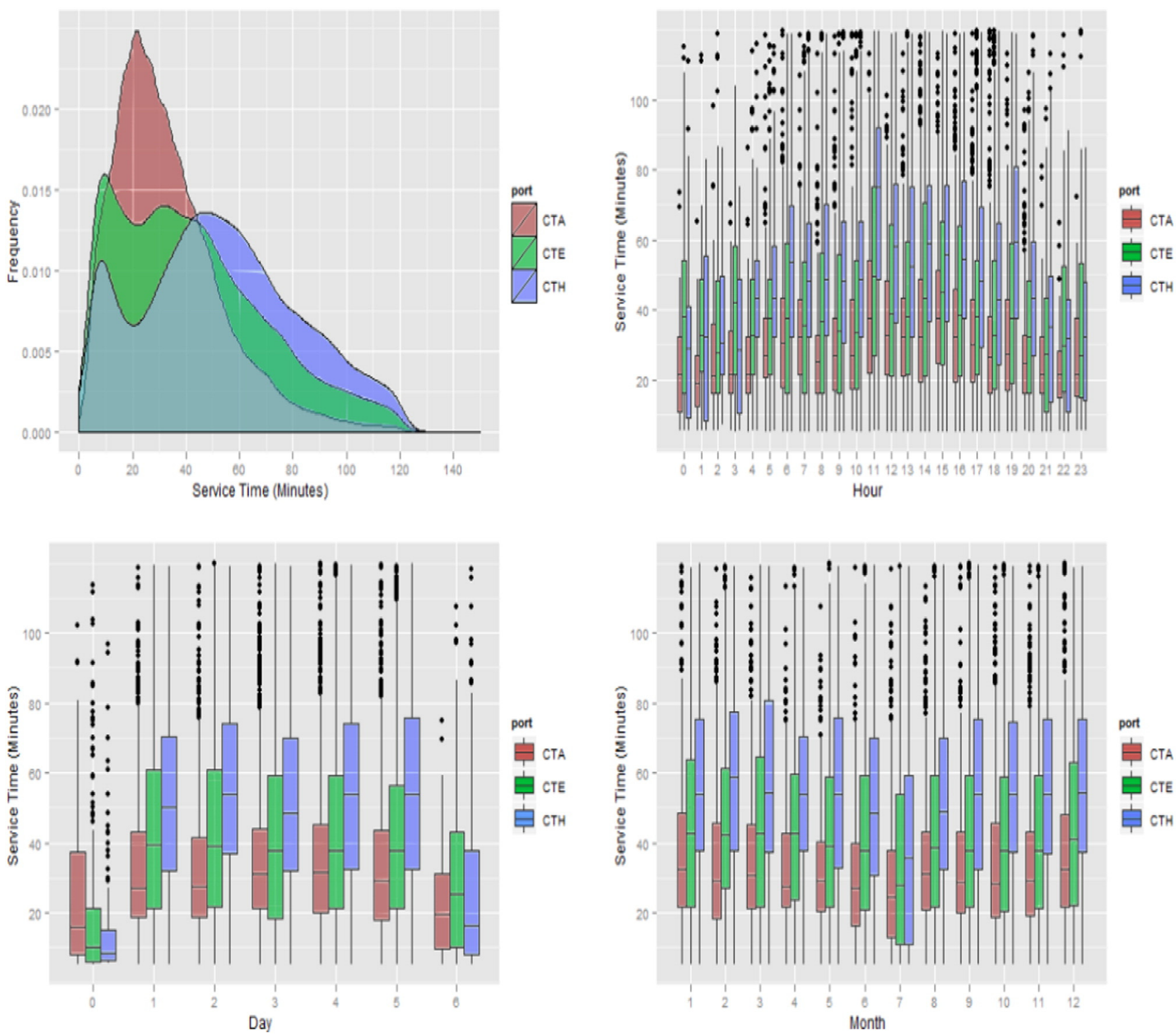

Fig. 5. The service rate profile of the respondent seaports.

Please cite this article as: M. Wasesa, et al., The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates, Decision Support Systems (2016), http://dx.doi.org/10.1016/j.dss.2016.11.008

Table 6
The performance of the prediction models.

Seaport		CTA				CTE				CTH			
Update rate		15 min		7.5 min		15 min		7.5 min		15 min		7.5 min	
Dataset		TR	TS	TR	TS	TR	TS	TR	TS	TR	TS	TR	TS
Mean*		31.16	31.16	31.82	31.76	39.36	39.29	39.84	39.86	50.83	50.80	51.05	51.10
St. Dev.*		15.38	15.33	15.27	15.12	19.73	19.71	20.16	20.15	22.26	22.42	23.02	23.14
Prediction models performance (RMSE)*	LM_0	14.93	14.89	14.97	14.83	19.30	19.29	20.01	20.03	21.15	21.25	22.60	22.77
	GBM_0	14.81	14.78	14.87	14.47	19.23	19.22	19.89	19.92	21.24	21.36	22.48	22.64
	LM_1	11.24	11.29	9.04	9.08	17.17	17.17	15.59	15.57	18.57	18.74	18.78	18.89
	GBM_1	10.45	10.56	8.23	8.25	16.40	16.44	14.98	14.99	17.73	17.91	17.33	17.50
	LM_2	11.18	11.23	8.97	9.06	17.06	17.08	15.54	15.53	18.46	18.63	18.74	18.86
	GBM_2	10.39	10.55	8.17	8.22	16.26	16.37	14.85	14.88	17.60	17.85	17.22	17.39
	LM_3	11.15	11.19	8.97	9.05	17.00	17.02	15.52	15.51	18.39	18.57	18.71	18.81
	GBM_3	10.38	10.52	8.12	8.17	16.25	16.35	14.79	14.83	17.59	17.84	17.18	17.34
	LM_4	11.14	11.16	8.97	9.05	16.96	16.98	15.50	15.49	18.34	18.51	18.68	18.79
	GBM_4	10.38	10.52	8.09	8.14	16.22	16.33	14.74	14.79	17.55	17.80	17.11	17.26
	LM_5	11.12	11.16	8.97	9.05	16.94	16.95	15.49	15.48	18.31	18.50	18.65	18.77
	GBM_5	10.35	10.51	8.08	8.13	16.22	16.34	14.73	14.80	17.50	17.74	17.08	17.26

TR = training dataset; TS = testing dataset; *RMSE, mean, and standard deviation in minutes.

(CTA = 15.38 min compared to CTE = 19.73 min and CTH = 22.26 min). The low standard deviation figures indicate the seaport's consistency in conducting the containers loading/unloading service.

Analyzing the prediction models' performance, we observe that including inertia effects in both the gradient boosting models (GBM_{1-5}) and the linear models (LM_{1-5}) improved the prediction performance significantly (lower RMSE value). The biggest prediction improvement was achieved when we added the inertia effects for the first time (GBM_1, LM_1). For instance, the prediction performance of the GBM model for the CTA seaport improved from 14.81 min ($GBM_{0\text{-}CTA}$) down to 10.45 min ($GBM_{1\text{-}CTA}$) (see Table 6). As more inertia effects were added, the prediction error decreased correspondingly. Fig. 6 highlights the prediction improvements for the 1500th quarter to the 2500th quarter. Note that the position of the prediction lines of GBM_1 and LM_1 are much closer to the yellow dots (the service rate actual value) compared to the prediction lines of GBM_0 and LM_0. This is valid for any model at any evaluated seaport (see Table 6). Note also that in general, the gradient boosting models performed better than the benchmark normal linear models.

Prediction models that were constructed for seaports with more consistent performance (smaller service rate standard deviation) performed better (lower RMSE) compared to models constructed for seaports with more volatile performance. Supporting the claim, we provide a snapshot of the seaports' actual service rate performance for the first week of July 2013 period in Fig. 7. As depicted, the CTA has more consistent service rate than its competitors and the CTH has the most volatile service rate. Note that the prediction models that were constructed for the CTA also have the best predictive performance (the lowest RMSE value) and the ones for the CTH also have the poorest performance (the highest RMSE value) (see Table 6).

Considering the high range of a seaport's service rate value that can reach up to 50 min (see Fig. 7), our solution can deliver reasonably good predictions. The RMSE value can go below 17.3 min for CTH and below 8.2 min for more consistent seaports (i.e. CTA). The findings confirm the usefulness and appropriateness of our system in supporting drayage operators in predicting a seaport's service rate so that truck route planning will minimize the time spent at stop points.

Fig. 6. The impact of inertia effects addition on prediction performance.

Please cite this article as: M. Wasesa, et al., The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates, Decision Support Systems (2016), http://dx.doi.org/10.1016/j.dss.2016.11.008

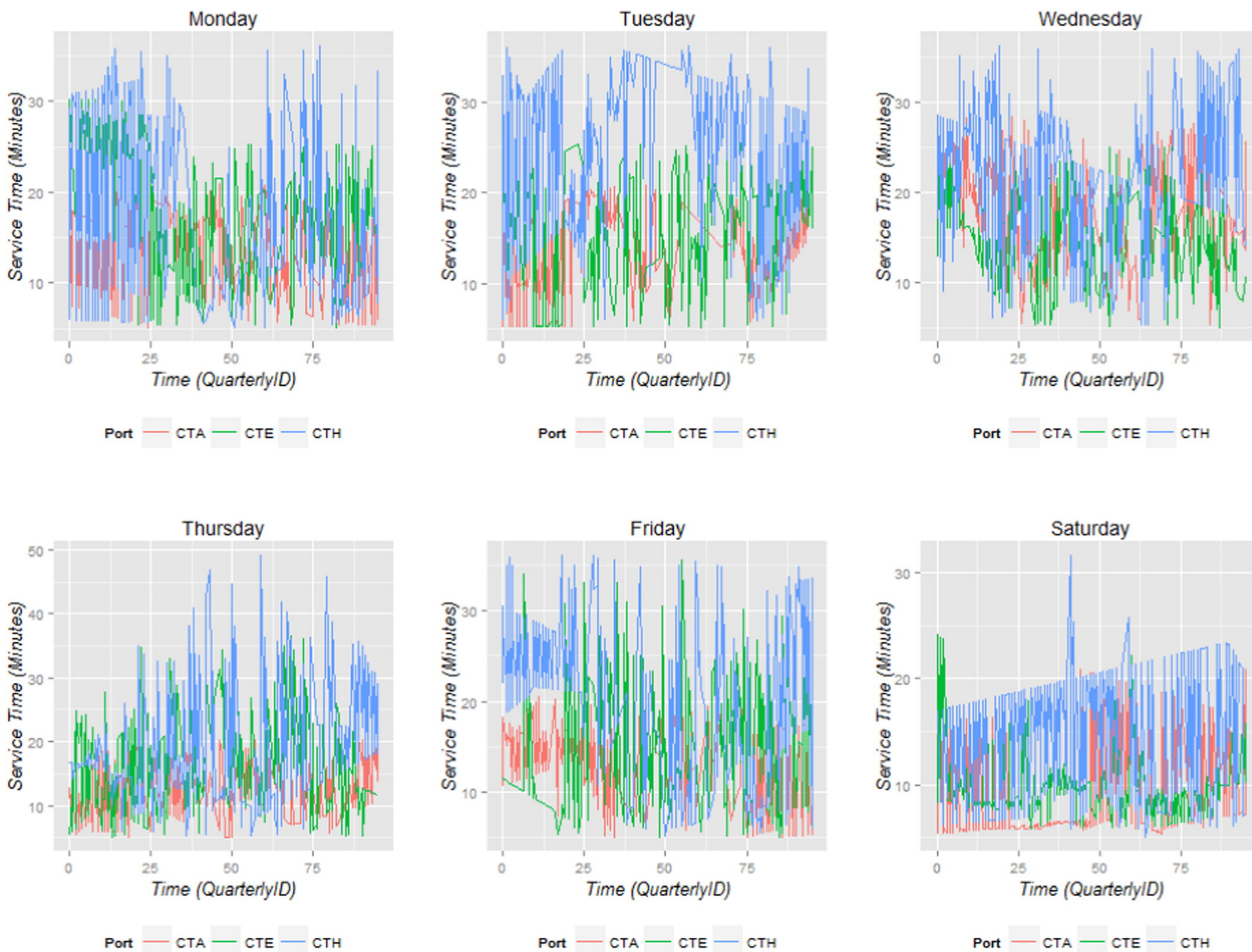

Fig. 7. The seaports' service rate performance for the July 2013 period.

6. Conclusions

With our seaport service rate prediction system, we provide a solution for predicting seaport service rate performance using drayage trucks' trajectory data. The proposed solution is constructed on three important components, namely trajectory reconstruction, geo-fencing analysis, and prediction model development. To validate the proposed prediction analytics solution, we analyzed more than 15 million mobility records logged from more than 200 trucks over a period of 19 months. Using a high volume of data with modest information features, we incorporate the temporal and the inertia effects as the main predictors. As the final result, the proposed gradient boosting model-based solution provides better predictions than the linear model benchmark solution.

From a practical point of view, the system can support drayage operators in predicting a seaport's service rate so that truck route planning will minimize the time spent at stop points. Recall that the application of the system is not limited to predicting the service rate of seaports and can also be applied to predict any service station's service rate. In generating predictions, the system uses the vehicle telematics data logs (trajectory data), circumventing the need to modify the existing appointment system.

From a research point of view this design science study [63,64] corresponds to the largely unexplored field of predictive analytics development using geospatial sensor-based data [19,20,22]. This study offers a new approach to the seaport congestion issue and explains how stakeholders can use predictive analytics techniques on their data assets, especially vehicle telematics data, to extract better insights to improve their decision making [15,27,51,56,58,62]. To the best of our knowledge, this approach has not been introduced in the research literature on seaport diversion initiatives (that aim to divert the road commodity flow onto alternative channels, such as extended gate and dry port initiatives) [7–10], non-diversion initiatives (that focus on improving the working condition of the seaport itself, by extending the seaport gate's opening hours and improving the seaport gate appointment system) [7,11–13], and decision support systems on seaport hinterland operation topics [30–32,35–37,40,41]. Moreover, while many studies attempted to predict the seaports' productivity for the seaside using yearly or monthly statistics archives [43–47], this study focuses on the seaports' landside productivity using sensor-based trajectory dataset that is updated every few minutes.

This study is not without limitations, some of which open opportunities for further research. The first limitation comes from the nature of the dataset. In this study the updating rate threshold value of the trajectory dataset is limited to 7.5 min. Setting a lower threshold is not possible since it will filter out most of the available data records. Secondly, to produce accurate predictions, the proposed solution requires high volume trajectory data logs that are gathered from many respondent trucks. Small drayage operators with few drayage trucks may need longer time to gather adequate data. As an option, one can explore the possibility of sharing trajectory data with other operators. Thirdly, we use the trajectory data for both training and testing the prediction models.

Please cite this article as: M. Wasesa, et al., The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates, Decision Support Systems (2016), http://dx.doi.org/10.1016/j.dss.2016.11.008

Using other data sources (surveys of stakeholders, seaport internal datasets) to test the prediction results can give more accurate predictive (RMSE) values. Fourthly, this study aims to build a generalizable system that uses the limited specifications of truck telematics data, namely the trajectory data, as the basis for constructing the prediction models. Alternatively, one can construct better performing systems by using richer telematics data specifications and combine data sets of collaborating drayage operators. Next, we only focus on the gradient boosting model as an example of a robust predictive analytics method. However, one can attempt to develop better performing systems by applying alternative predictive models such as random forest, support vector regression, and so on [15,65].

Acknowledgment

This research is funded by the Erasmus Research Institute of Management (ERIM) of Erasmus University Rotterdam and Almende BV, The Netherlands.

Appendix A. Summary of notation

Symbol	Definition
i	Seaport's identity
x	Truck's identity
$X = \{x_1, x_2, \ldots x_n\}$	Set of respondent trucks
t	Time index
$\theta_{t_{arrivalx}}{}^{i}$	Time of arrival of respondent truck *x* at seaport *i*
$\theta_{t_{departurex}}{}^{i}$	Time of departure of respondent truck *x* at seaport *i*
$\varphi_{t_x}{}^{i}$	Service rate performance of seaport *i* at a specific time period *t* as measured by respondent truck *x*
c^i	User's estimation of the seaport *i* service rate value
$\overline{\varphi_t^i}$	Predicted value of the service rate of seaport *i* at time *t*
$\alpha_i(t)$	Temporal effect predictors
$\beta_i(\gamma_t^i)$	Inertia effect predictors
ε_{it}	Prediction error
$\dot{\varphi_{t-z}^{i}}$	Historical trace of the seaport *i* service rate, whereas $\dot{\varphi_{t-z}^{i}} = \{ \varphi_{t-z}^i, \ldots, \varphi_{t-2}^i, \varphi_{t-1}^i\}$
ζ_t^i	Number of arriving trucks at the seaport *i* at time *t*
$\dot{\zeta_{t-z}^{i}}$	Historical trace of the number of arriving trucks at seaport *i* at time *t*, whereas $\dot{\zeta_{t-z}^{i}} = \{ \zeta_{t-z-1}^i, \ldots, \zeta_{t-2}^i, \zeta_{t-1}^i\}$
δ_t^i	Number of departing trucks at the seaport *i* at time *t*
$\dot{\delta_{t-z}^{i}}$	Historical trace of the number of departing trucks at seaport *i* at time *t*, whereas $\dot{\delta_{t-z}^{i}} = \{\delta_{t-z-1}^i, \ldots, \delta_{t-2}^i, \delta_{t-1}^i\}$
z	Inertia effect index
$F_i(\psi_t)$	Prediction function
ψ_t	Potential predictors
$F_i^{(m)}(\psi_t)$	The estimation of $F_i(\psi_t)$ at the *m*-th iteration
m	Iteration index
$h_m(\psi_t, Ш_m)$	Weak learner estimate for the gradient boosting model
v	Shrinkage parameter for the gradient boosting model

References

[1] G. Chen, K. Govindan, Z. Yang, Managing truck arrivals with time windows to alleviate gate congestion at container terminals, International Journal of Production Economics 141 (2013) 179–188.
[2] M. Acciaro, A. Mckinnon, Efficient hinterland transport infrastructure and services for large container ports, International Transport Forum Discussion Paper 19 (2013) 1–32.
[3] N. Sathaye, R. Harley, S. Madanat, Unintended environmental impacts of nighttime freight logistics activities, Transportation Research Part A: Policy and Practice 44 (2010) 642–659.
[4] R. Midoro, E. Musso, F. Parola, Maritime liner shipping and the stevedoring industry: market structure and competition strategies, Maritime Policy & Management 32 (2005) 89–106.
[5] E. Ursavas, A decision support system for quayside operations in a container terminal, Decision Support Systems 59 (2014) 312–324.
[6] M.R. van der Horst, P.W. de Langen, Coordination in hinterland transport chains: a major challenge for the seaport community, Maritime Economics and Logistics 10 (2008) 108–129.
[7] A. Maguire, S. Ivey, M.M. Golias, M.E. Lipinski, Relieving congestion at intermodal marine container terminals: review of tactical/operational strategies, Proc. 51st Annu. Transp. Res. Forum 2010, pp. 631–645.
[8] A. Veenstra, R. Zuidwijk, E. van Asperen, The extended gate concept for container terminals: expanding the notion of dry ports, Maritime Economics and Logistics 14 (2012) 14–32.
[9] V. Roso, K. Lumsden, A review of dry ports, Maritime Economics and Logistics 12 (2010) 196–213.
[10] K. Cullinane, R. Bergqvist, G. Wilmsmeier, The dry port concept – theory and practice, Maritime Economics and Logistics 14 (2012) 1–13.
[11] G. Giuliano, T.O. Brien, S. Hayden, P. Dell, T.O. Brien, The terminal gate appointment system at the ports of Los Angeles and long beach: an assessment, METRANS Rep., 1, 2008, pp. 1–30.
[12] G. Giuliano, T. O'Brien, Reducing port related truck emmisions: the terminal gate appointment system at the ports of Los Angeles and long beach, Transportation Research Part D 12 (2007) 460–473.
[13] P. Morais, E. Lord, Terminal appointment system study, Transportation Research Board 1 (2006) 1–126.
[14] J. Larsen, SUPPORT - a proposal to strengthening port of Rotterdam's logistics network, Almende Research 1 (2009) 1–16.
[15] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, second ed. Springer Science Business Media, LLC, 2011.
[16] M. Baumgartner, J. Léonardi, O. Krusch, Improving computerized routing and scheduling and vehicle telematics: a qualitative survey, Transportation Research Part D: Transport and Environment 13 (2008) 377–382.
[17] T.G. Crainic, M. Gendreau, J.-Y. Potvin, Intelligent freight-transportation systems: assessment and the contribution of operations research, Transportation Research Part C: Emerging Technologies 17 (2009) 541–557.
[18] G. Shmueli, O.R. Koppius, Predictive analytics in information systems research, MIS Quarterly 35 (2011) 553–572.
[19] H.J. Watson, B.H. Wixom, The current state of business intelligence, IEEE Computer 40 (2007) 96–99.
[20] H. Chen, R.H.L. Chiang, V.C. Storey, Business intelligence and analytics: from big data to big impact, MIS Quarterly 36 (2012) 1165–1188.
[21] G. Kortuem, F. Kawsar, V. Sundramoorthy, D. Fitton, Smart objects as building blocks for the internet of things, IEEE Internet Computing 14 (2010) 44–51.
[22] S. Negash, Business intelligence, Communications of the Association for Information Systems 13 (2004) 177–195.
[23] S.T. March, A.R. Hevner, Integrated decision support systems: a data warehousing perspective, Decision Support Systems 43 (2007) 1031–1043.
[24] N. Eagle, A.S. Pentland, Eigen behaviors: identifying structure in routine, Behavioral Ecology and Sociobiology 63 (2009) 1057–1066.
[25] N. Uno, F. Kurauchi, H. Tamura, Y. Iida, Using bus probe data for analysis of travel time variability, Journal of Intelligent Transportation Systems 13 (2009) 2–15.
[26] Y.U. Zheng, Trajectory data Mining : an overview, ACM Transactions on Intelligent Systems and Technology 6 (2015) 1–41.
[27] J.C. Herrera, D.B. Work, R. Herring, X.(J.). Ban, Q. Jacobson, A.M. Bayen, Evaluation of traffic data obtained via GPS-enabled mobile phones: the mobile century field experiment, Transportation Research Part C: Emerging Technologies 18 (2010) 568–583.
[28] G. Hiribarren, J.C. Herrera, Real time traffic states estimation on arterials based on trajectory data, Transportation Research Part B: Methodological 69 (2014) 19–30.
[29] Z. Sun, X.(J.). Ban, Vehicle classification using GPS data, Transportation Research Part C: Emerging Technologies 37 (2013) 102–117.
[30] N. Huynh, M. Walton, Improving efficiency of drayage operations at seaport container terminals through the use of an appointment system, in: J. Boese (Ed.), Handb. Termin. Plan, Operations, Springer Science Business Media, LLC, Hamburg, Germany 2011, pp. 323–344.
[31] R. Namboothiri, A.L. Erera, Planning local container drayage operations given a port access appointment system, Transportation Research Part E 44 (2008) 185–202.
[32] Z.-H. Hu, Z.-H. Sheng, A decision support system for public logistics information service management and optimization, Decision Support Systems 59 (2014) 219–229.
[33] N. Li, G. Chen, K. Govindan, Z. Jin, Disruption management for truck appointment system at a container terminal: a green initiative, Transportation Research Part D: Transport and Environment 1 (2016) 1–13.
[34] M.-H. Phan, K.H. Kim, Negotiating truck arrival times among trucking companies and container terminal, Transportation Research Part E: Logistics and Transportation Review 75 (2015) 132–144.
[35] D.L. Bandeira, J.L. Becker, D. Borenstein, A DSS for integrated distribution of empty and full containers, Decision Support Systems 47 (2009) 383–397.
[36] K.G. Murty, J. Liu, Y. Wan, R. Linn, A decision support system for operations in a container terminal, Decision Support Systems 39 (2005) 309–332.
[37] N.N. Huynh, C.M. Walton, R. River, Methodologies for reducing truck turn time at marine container terminals, Center for Transportation Research - The University of Texas at Austin 7 (2005) 1–144.
[38] E. Zehendner, D. Feillet, Benefits of a truck appointment system on the service quality of inland transport modes at a multimodal container terminal, European Journal of Operational Research 235 (2014) 461–469.
[39] C. Guan, R. Liu, Container terminal gate appointment system optimization, Maritime Economics and Logistics 11 (2009) 378–398.
[40] P. Ioannou, A. Chassiakos, H. Jula, G. Valencia, Cooperative time window generation for cargo delivery/pick up with application to container terminals, METRANS Rep 2005, pp. 3–18.
[41] W. Zhao, A.V. Goodchild, The impact of truck arrival information on container terminal rehandling, Transportation Research Part E: Logistics and Transportation Review 46 (2010) 327–343.

Please cite this article as: M. Wasesa, et al., The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates, Decision Support Systems (2016), http://dx.doi.org/10.1016/j.dss.2016.11.008

[42] J. Tongzon, Efficiency measurement of selected Australian and other international ports using data envelopment analysis, Transportation Research Part A: Policy and Practice 35 (2001) 107–122.
[43] K. Fung, Competition between the ports of Hong Kong and Singapore: a structural vector error correction model to forecast the demand for container handling services, Maritime Policy & Management 28 (2001) 3–22.
[44] K. Cullinane, T. Wang, D. Song, P. Ji, The technical efficiency of container ports: comparing data envelopment analysis and stochastic frontier analysis, Transportation Research Part A: Policy and Practice 40 (2006) 354–374.
[45] S.H. Chen, J.N. Chen, Forecasting container throughputs at ports using genetic programming, Expert Systems with Applications 37 (2010) 2054–2058.
[46] J. Geng, M.W. Li, Z.H. Dong, Y.S. Liao, Port throughput forecasting by MARS-RSVR with chaotic simulated annealing particle swarm optimization algorithm, Neurocomputing 147 (2015) 239–250.
[47] UNECE, Introducing UN/EDIFACT, United Nations Rules Elec-Tronic Data Interchang. Adm. Commer. Transp, 2016 1 (http://www.unece.org/cefact/edifact/welcome.html (accessed January 1, 2016)).
[48] O. Braysy, M. Gendreau, Vehicle routing problem with time windows, part I: route construction and local search algorithms, Transportation Science 39 (2005) 104–118.
[49] A.L. Erera, J.C. Morales, M. Savelsbergh, The vehicle routing problem with stochastic demand and duration constraints, Transportation Science 44 (2010) 474–492.
[50] A. Gunasekaran, B. Kobu, Performance measures and metrics in logistics and supply chain management: a review of recent literature (1995-2004) for research and applications, International Journal of Production Research 45 (2007) 2819–2840.
[51] D.B. Work, O.-P. Tossavainen, S. Blandin, A.M. Bayen, T. Iwuchukwu, K. Tracton, An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices, IEEE Conf. Decis. Control, IEEE 2008, pp. 5062–5068.
[52] A. Sheth, S. Seshan, D. Wetherall, Geo-fencing: confining Wi-Fi coverage to physical boundaries, Pervasive Computing 1 (2009) 274–290.
[53] T. Hastie, R. Tibshirani, Generalized additive models, Statistical Science 1 (1986) 297–318.
[54] S. Ben Taieb, R.J. Hyndman, A gradient boosting approach to the Kaggle load forecasting competition, International Journal of Forecasting 1 (2013) 1–13.
[55] V. Cantillo, J.D. Ortuzar, H. Williams, Modeling discrete choices in the presence of inertia and serial correlation, Transportation Science 41 (2007) 195–205.
[56] J.H. Friedman, Stochastic gradient boosting, Computational Statistics and Data Analysis 38 (2002) 367–378.
[57] J.H. Friedman, Greedy function approximation: a gradient boosting machine, The Annals of Statistics 29 (2001) 1189–1232.
[58] T.G. Dietterich, Ensemble methods in machine learning, Multiple Classifier Systems (1990) 1–15.
[59] R.E. Schapire, The boosting approach to machine learning: an overview, Nonlinear Estimation and Classification 171 (2003) 149–171.
[60] D. Mease, Evidence contrary to the statistical view of boosting: a rejoinder to responses, Journal of Machine Learning Research 9 (2008) 195–201.
[61] Port of Rotterdam, Facts & figures about the port, Port Rotterdam off. website, 2016 1 (https://www.portofrotterdam.com/en/the-port/facts-figures-about-the-port (accessed June 26, 2016)).
[62] G.A. Giannopoulos, Towards a European ITS for freight transport and logistics: results of current EU funded research and prospects for the future, European Transport Research Review 1 (2009) 147–161.
[63] S.T. March, G.F. Smith, Design and natural science research on information technology, Decision Support Systems 15 (1995) 251–266.
[64] S. Gregor, A.R. Hevner, Positioning and presenting design science research for maximum impact, MIS Quarterly 37 (2013) 337–355.
[65] T.M. Mitchell, Machine Learning, McGraw-Hill, 1997.

Meditya Wasesa is the co-founder and director of The Indonesian Center for Logistics and Value Chains (ICLOV). He earned a bachelor degree in mechanical engineering from Bandung Institute of Technology, Indonesia and a master degree (cum laude) in logistics engineering from the University of Duisburg-Essen, Germany. During his time in Germany, he had also worked for General Motors Europe GmbH. He is currently working on a PhD research at the Rotterdam School of Management of Erasmus University funded by the Erasmus Research Institute of Management and Almende BV, the Netherlands. His research focuses on the design, role, and impact of agent-based inter-organizational systems, enterprise information systems, business intelligence, and predictive analytics in advanced logistics operations.

Andries Stam (CEO Almende BV) received his MSc in Computer Science at the University of Leiden and his PhD at the same university on a thesis about the modeling of interaction in distributed evolving software systems. Within Almende, Andries coordinates all R&D activities of the company. He has been involved in more than 40 research projects (both National and European) in various domains: fundamentals of software engineering, embedded systems, robotics, logistics, security, care & wellbeing and energy. The contribution of Almende in all these fields revolves around the idea of self-organisation and combines essential insights into four knowledge domains: Internet of Things, Agent Technology, Bio-inspired Systems and Visual Analytics. Andries has published in the areas of software engineering, coordination languages, and enterprise architecture.

Eric van Heck is Professor of Information Management and Markets and Chairman of the Department of Technology & Operations Management at the Rotterdam School of Management of Erasmus University. His research concentrates on the role and impact of business architectures and digital platforms to tackle complex societal and business challenges. At the moment he is working on digital business and architecture, big data and analytics, digital work, and energy markets. His research is carried out in collaboration with innovative companies and universities in Brazil, China, Europe, Indonesia, and the United States.

Please cite this article as: M. Wasesa, et al., The seaport service rate prediction system: Using drayage truck trajectory data to predict seaport service rates, Decision Support Systems (2016), http://dx.doi.org/10.1016/j.dss.2016.11.008