ARTICLE TYPE

Predicting Ship Arrival Times at (Provincial) Bridges

Michael de Winter,[1] Bram Albers,[2] and Eric Wessels[2]

[1]I&A division, Provincie Zuid Holland
[2]Analyze, KPenP
Author for correspondence: M.R. de Winter, Email: mr.de.winter@pzh.nl.

Abstract

This paper proposes a method to predict ship arrival times, specifically measured 500 meters before reaching provincial bridges. Our approach leverages a machine learning model trained on three years of Automatic Identification System (AIS) data. This model integrates historical arrival times and various real-time factors, such as the current speed of the ship.
Utilizing a random forest regression model, we achieved a root mean square error (RMSE) of approximately 5 minutes between predicted and actual ship arrival times, as validated by the three-year data set split into a training and test set. However, the accuracy varied slightly depending on the specific bridge and the distance to the bridge, with shorter distances being more accurate to predict correct ETA's.
The precise prediction of ship arrival times could yield considerable advantages for traffic management. Such foresight permits proactive traffic redirection to faster routes, which is particularly critical for emergency vessels like ambulances and fire boats.
In addition, accurate ship arrival time predictions aid in synchronizing the transit of multiple vessels during a single bridge lift. This is particularly beneficial when these vessels are scheduled to arrive in close succession. These are just a few examples of the myriad benefits that precise ship arrival forecasts offer for efficient maritime traffic management.
Moreover, the model contributes to environmental sustainability. By reducing traffic congestion, it indirectly leads to lower greenhouse gas emissions, emphasizing the model's climate-positive potential.

Keywords: Ship ETA prediction, Traffic management, CO2 reduction, Congestion reduction

1. Introduction

The Province of South Holland manages several bridges, especially in dense packed region of South Holland a bridge opening will certainly cause congestion especially during rush hours CBS 2010.

This issue has partially successfully elicited calls for action from several political parties, aiming to address the traffic congestion resulting from bridge openings. For instance now only for commercial shipping's are bridge openings allowed during rush hours Zuid-Holland, n.d.

Although it is not feasible to entirely eliminate congestion caused by bridge operations, it is possible to optimize it using various strategies **sspzh2019**.

One such strategy involves predicting the estimated times of arrival (ETAs) for ships at bridges. With accurate ETAs, road traffic could be rerouted to potentially faster alternatives, or bridge openings can be consolidated for multiple ships if their arrival times are close together.

This paper primarily focuses on predicting a ship's ETA 500 meters before it reaches a bridge. This distance is significant because bridge operators commence the bridge opening process when ships are 500 meters away.

The training data for our model consists of three years of historical Automatic Identification System (AIS) data. In brief, AIS data is gathered from transceivers mounted on ships, which provide unique identification for each vessel, GPS coordinates (longitude and latitude), the ship's course, and speed. This data is acquired from Rijkswaterstaat (RWS) Vaarweginformatie, n.d.

2. Related work

Bodunov et al. 2018 presented a study that not only predicts a ship's ETA but also its anticipated destination. They employed a feed-forward neural network model for these predictions. However, the paper appears somewhat limited in detailing the variables incorporated into the model, focusing primarily on the architecture model itself. Their work primarily aims at predicting events weeks ahead, in contrast to our focus on minute-scale predictions.

Veenstra and Harmelink 2021 et al explored the use of previous ETA predictions as inputs for ETA prediction models. Their findings demonstrate that previous ETA predictions hold predictive value, likely due to the model's ability to adjust and learn from prior erroneous predictions, exhibiting a form of self-learning.

Flapper 2020 et al employed gradient boosting to rank the importance of various features in predicting ETAs. They identified the following order of importance for the features: Distance to ETA target, Latitude of ETA target, Longitude of ETA target, Target ETA ID, Time from previous point, Ship Length, Current Longitude, Previous ID, Distance from previous point, Ship Width, Previous Longitude, Current Latitude, Previous Latitude, Ship Depth, Day, Current ID, Hour, Ship Type, and Month.

In their work at Uber, Hu et al. 2022 replaced an XGBoost algorithm with a deep learning residual model for ETA prediction. This model utilizes the distance and maximum driving speed for a selected route from a routing system as inputs. It then attempts to adjust the predictions by incorporating residual parameters based on real-time data.

3. Materials and Methods

3.1 Study area

We focused solely on provincial waterways, which means that only AIS data corresponding to these waterways were included in our analysis. Additionally, we considered only the provincial bridges in the vicinity of Leiden, as illustrated in the study area displayed Figure 1.

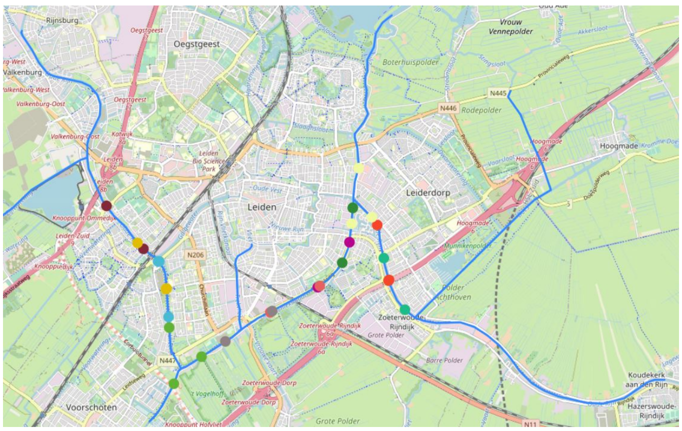

Figure 1. The study area encompasses provincial waterways, highlighted in blue, and ETA prediction points for each bridge in the region if Leiden. The color of each circle represents a bridge, which may have one or multiple sides for making ETA predictions.

3.2 Data preprocessing

Data preprocessing is necessary since the raw AIS data must be transformed to determine whether specific ships pass certain bridges. Additionally, the provincial waterways are segmented into 100-meter sections, which simplifies the calculation of ETAs based on AIS positions.

3.2.1 Bridge Passages

We have three years of AIS data, comprising records from 2019, 2020, and 2021. To determine whether a particular vessel, as identified by its AIS trace data, will pass under a given bridge, we must preprocess the data to calculate potential bridge passages. To achieve this, we draw a line across each bridge and then verify whether the AIS trace data of a specific ship passes this line, on either the right or left side. As shown in Figure 2.

3.2.2 Provincial Waterway 100 meter Sections

To streamline the process, we divide the provincial waterways into sections of 100 meters each. This strategy simplifies ETA predictions as it shifts the focus from calculating an ETA for a specific AIS GPS position to generating predictions for each 100-meter section of the provincial waterways for each bridge. This approach also yields more data for making predictions. Gathering a sufficient amount of data for a specific AIS GPS position is considerably more challenging than for a 100-meter waterway section. But even with 100 meter sections, there

Figure 2. In order to see if a AIS trace passages a certain bridge we draw a line to see if the AIS data passes this line.

are still some sections in the provincial waterways which don't contain any data, we copied data from nearby segments if this occured.

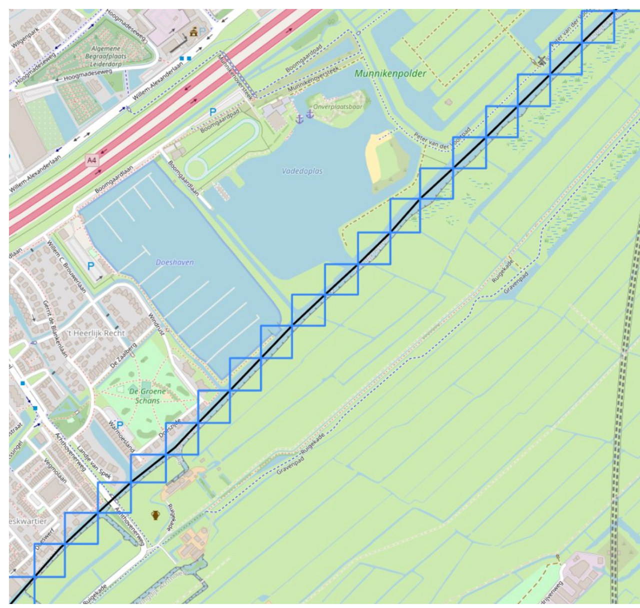

Figure 3. The provincial waterways are divided into 100 meter sections, for each of these sections the average historical arrival times are calculated to various bridges.

For each of these 100-meter sections of provincial waterways, we calculate the average historical travel time taken by ships to reach a given 500 meter point for a given bridge more on this later. We make the assumption that a ship takes the shortest route to a specific bridge. To ascertain this shortest route, we represent the sections as a graph and apply a shortest path algorithm.

We then store each section and its associated average historical travel time to each bridge in a database. With this setup, a simple database lookup can be performed for a specific AIS GPS position. This lookup will determine which section of the provincial waterway it is in and reveal the historical travel times to reach a certain bridge. This forms one component of our prediction model but can already be utilized to predict ETA's.

4. Input Data and Model

Inspired by Hu et al. 2022 we combine the average historical travel time in a Provincial Waterway segments, which can be seen a basic routing system, to a bridge with real time information variables. The input variables for the model are shown in Table 1

Table 1. Input variables with the best performance for the model. *Section 4.1 will go deeper into this.

Variables
Historical Travel Time *
Current Speed Over Ground
Category Ship
Randomised Shipid
Length Ship
Current Month

These variables were all picked because of there best predictive performance, which was tested on the training set.

4.1 Historical Average Travel Time to Bridge

We calculate the historical average travel time to a bridge as follows. We start with the assumption that a ship located in the current provincial waterway segment will take the shortest path to a bridge. To determine this route and its associated distance, we represent the segments as a graph and apply a shortest path algorithm. We then divide this distance by 1.94444. This factor is equivalent to a speed of 7 km per hour we, the typical speed at which ships sail we calculated from the data, but converted into seconds. As a result, we obtain the estimated number of seconds it would take for the ship to reach the specific bridge. This value thus represents the historical average travel time to a bridge.

4.2 RandomForest Regressor model

We employ the Random Forest Regressor model provided by the Python scikit-learn package Pedregosa et al. 2011. In short a random forest Regressor model makes multiple decisions tree than average there outcomes for making predictions.

To find the optimal parameters, we performed a grid search, which identified the optimal parameters as a 'random_state' of 42 and the 'criterion' set to MSE. The data was divided into training and testing datasets, with a 70/30 percent split favoring training data over testing data, respectively.

We have experiment with one Random Forest Regressor model for all bridges and one Random Forest Regressor model for each specific bridge. Greater accuracy was achieved with one Random Forest Regressor model per bridge and as such one Random Forest Regressor model was made per bridge.

4.3 Results

The table below presents the results in terms of Mean Error and Root Mean Squared Error (RMSE), quantifying the difference between the actual arrival time and the predicted ETA 500 meters before a bridge. Given the potential discrepancy between the 500-meter mark before a bridge and the nearest AIS signal to that location, these results should be interpreted as indicative rather than definitive. The results can differ depending on the distance to the bridge, in the appendix Table 3 shows error message on differing distances.

Table 2. Results on model predictions difference between the ETA and the actual arrival time before a bridge on the test set.

Bridge	Mean Error	RMSE
Wilhelminabrug Leiden	0:02:50	0:04:53
Julius Caesarbrug	0:02:55	0:04:44
Leiderdorpsebrug	0:02:41	0:05:06
Waddingerbrug	0:03:22	0:05:52
Rhijnvreugdbrug	0:03:01	0:05:42
Kanaalbrug	0:02:56	0:05:11
Lammebrug	0:03:40	0:05:58
Hooghkamerbrug	0:03:07	0:05:09
Spanjaardsbrug	0:03:43	0:06:44
Stevensbrug	0:03:28	0:05:58
Hoflandbrug	0:03:05	0:05:31

5. Discussion

RME for all the bridges was around 5 minutes on the test set, if this error margin is acceptable is still being internally debated. Error margin here can be crucial for adoption. If the error margin is too large trust might not be lost in the model.

Further enhancements can be potentially achieved through the implementation of deep learning models, however, such investigation extends beyond the scope of this current research. As suggested by Veenstra (2021), there may be significant merit in conceptualizing the data as a time series. This could potentially yield meaningful insights as past speed and estimated time of arrival (ETA) predictions could harbor predictive capacity.

References

Bodunov, Oleh, Florian Schmidt, André Martin, Andrey Brito, and Christof Fetzer. 2018. Real-time destination and eta prediction for maritime traffic. In *Proceedings of the 12th acm international conference on distributed and event-based systems,* 198–201.

CBS. 2010. Kerncijfers provinciemonitor; 2006 - 2009. Accessed April 3, 2010. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/71595ned/table.

Flapper, Edwin. 2020. Eta prediction for vessels using machine learning. B.S. thesis, University of Twente.

Hu, Xinyu, Tanmay Binaykiya, Eric Frank, and Olcay Cirit. 2022. Deepreta: an eta post-processing system at scale. *arXiv preprint arXiv:2206.02127.*

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Vaarweginformatie. n.d. https://vaarweginf ormatie.nl/f rp/main/#/home.

Veenstra, Albert, and Rogier Harmelink. 2021. On the quality of ship arrival predictions. *Maritime Economics & Logistics,* 1–19.

Zuid-Holland, Provincie. n.d. Brugopeningen en bedientijden. https://www.zuid-holland.nl/onderwerpen/verkeer-vervoer/vaarwegen/brugopeningen/.

Appendix 1. Appendix

Table 3. Results on model predictions difference between the ETA and the actual arrival time before a bridge on the test set on different times until bridge arrival

Bridge	Mean Error	RMSE	Time until bridge
Wilhelminabrug Leiden	0:02:04	0:03:17	0-15 minutes
Wilhelminabrug Leiden	0:02:59	0:04:38	25-30 minutes
Wilhelminabrug Leiden	0:03:01	0:04:37	40-45 minutes
Wilhelminabrug Leiden	0:08:13	0:10:55	55-60 minutes
Julius Caesarbrug	0:02:29	0:04:00	0-15 minutes
Julius Caesarbrug	0:03:17	0:05:01	25-30 minutes
Julius Caesarbrug	0:03:06	0:04:50	40-45 minutes
Julius Caesarbrug	0:05:54	0:08:35	55-60 minutes
Leiderdorpsebrug	0:01:54	0:03:21	0-15 minutes
Leiderdorpsebrug	0:02:46	0:04:38	25-30 minutes
Leiderdorpsebrug	0:05:00	0:07:09	40-45 minutes
Leiderdorpsebrug	0:10:41	0:15:43	55-60 minutes
Waddingerbrug	0:02:03	0:03:21	0-15 minutes
Waddingerbrug	0:03:02	0:04:36	25-30 minutes
Waddingerbrug	0:03:30	0:05:25	40-45 minutes
Waddingerbrug	0:06:28	0:10:37	55-60 minutes
Rhijnvreugdbrug	0:01:53	0:03:16	0-15 minutes
Rhijnvreugdbrug	0:02:34	0:04:06	25-30 minutes
Rhijnvreugdbrug	0:06:09	0:09:35	40-45 minutes
Rhijnvreugdbrug	0:08:43	0:13:22	55-60 minutes
Kanaalbrug, Leiden	0:02:03	0:03:32	0-15 minutes
Kanaalbrug, Leiden	0:03:23	0:05:14	25-30 minutes
Kanaalbrug, Leiden	0:03:16	0:04:53	40-45 minutes
Kanaalbrug, Leiden	0:05:36	0:09:10	55-60 minutes
Lammebrug	0:03:04	0:05:22	0-15 minutes
Lammebrug	0:03:55	0:05:26	25-30 minutes
Lammebrug	0:04:21	0:06:15	40-45 minutes
Lammebrug	0:06:06	0:09:27	55-60 minutes
Hooghkamerbrug	0:02:07	0:04:11	0-15 minutes
Hooghkamerbrug	0:03:06	0:04:14	25-30 minutes
Hooghkamerbrug	0:03:49	0:05:42	40-45 minutes
Hooghkamerbrug	0:05:46	0:08:18	55-60 minutes
Spanjaardsbrug	0:02:20	0:03:46	0-15 minutes
Spanjaardsbrug	0:03:28	0:05:01	25-30 minutes
Spanjaardsbrug	0:06:03	0:08:58	40-45 minutes
Spanjaardsbrug	0:12:09	0:16:58	55-60 minutes
Stevensbrug	0:02:28	0:04:10	0-15 minutes
Stevensbrug	0:03:31	0:05:03	25-30 minutes
Stevensbrug	0:03:44	0:05:35	40-45 minutes
Stevensbrug	0:08:11	0:13:07	55-60 minutes
Hoflandbrug	0:02:21	0:04:41	0-15 minutes
Hoflandbrug	0:02:53	0:04:29	25-30 minutes
Hoflandbrug	0:03:30	0:04:54	40-45 minutes
Hoflandbrug	0:06:49	0:11:28	55-60 minutes