Modelling of an Efficient System for Predicting Ships' Estimated Time of Arrival

by

Sadia Tasneem Rahman (70021105)

Md. Raqibur Rahman (170021115)

Ehtashamul Haque (170021130)

BACHELOR OF SCIENCE

IN

ELECTRICAL AND ELECTRONIC ENGINEERING

Department of Electrical and Electronic Engineering

Islamic University of Technology (IUT)

Board Bazar, Gazipur-1704, Bangladesh.

November, 2017.

CERTIFICATE OF APPROVAL

The thesis titled, "**Modelling of an Efficient System for Predicting Ships' Estimated Time of Arrival**" accepted as partial fulfillment of the requirement for the Degree of BACHELOR OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING of Islamic University of Technology (IUT).

Approved by:

Dr. Khondokar Habibul Kabir
(Supervisor)
Professor,
Electrical and Electronic Engineering Department,
Islamic University of Technology (IUT), Gazipur.

This thesis must be modified and re-submitted.

Declaration of Candidate

It is hereby declared that this thesis report is only submitted to The Electrical and Electronic Engineering Department any part of it has not been submitted elsewhere for the award of any Degree or Diploma.

Sadia Tasneem

Sadia Tasneem Rahman
Student ID: 170021105
Date: 10. 05. 2022

Raqibur Rahman

Md. Raqibur Rahman
Student ID: 170021115
Date: 9.05. 2022

Ehtashamul Haque

Ehtashamul Haque
Student ID: 170021130
Date: 10.5.2022

Dedicated to our conscience

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIS	Automatic Identification System
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
COG	Course Over Ground
CSV	Comma Seperated Value
ETA	Estimated Time of Arrival
ETD	Estimated Time of Departure
GPS	Global Positioning System
IMO	International Maritime Organization
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MDP	Markov Decision Process
ML	Machine Learning
MMSI	Maritime Mobile Service Identity
NMEA	National Marine Electronics Association
ReLU	Rectified Linear Unit
RMSE	Root Mean Squared Error
RL	Reinforcement Learning
SOG	Speed Over Ground
SVM	Support Vector Machine
UTC	Universal Time Coordinated

Acknowledgment

Alhamdulillahi Rabbil Alamin. At first, we praise and glorify the Lord, Allah Subhanahu wa ta'ala, as He is ought to be praised and glorified.

We would like to express our sincere and heartfelt gratitude to our supervisor, Dr. Khondokar Habibul Kabir, Professor, Department of Electrical and Electronic Engineering, Islamic University of Technology, for kindly supervising our thesis. Thanks a lot to sir for providing us with proper guidelines and keeping us focused on our objectives even during this challenging time of the COVID 19 pandemic. Without his affectionate supervision, this work might not have been completed. We are grateful to him for the effort and support he has given throughout the course of our thesis.

Then, I would like to express my gratitude to Dr. Yaseen Adnan Ahmed, Yaseen Adnan Ahmed, Research Associate, Maritime Safety Research Centre, Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde for his kind help, support, and guidance during the concept development and for monitoring the progress of the study work.

We are grateful to Raihan Seraj vai for providing us with valuable suggestions to continue our research work.

We are grateful to our fellow classmates of IUT for their continuous technical and mental support throughout the journey.

Also, we would like to thank our parents for their continuous support. Without them, it would have become difficult for us to finish our work. Lastly, we would like to thank all our well-wishers.

Abstract

Ports serve as a focal point for the global economy. Around 80% of global trade logistics is carried out through ports. As such, port efficiency is an important factor that has to be maintained properly in order to ensure a maximum economic gain. In order to boost port efficiency, a smart port system integrates state-of-the-art technology with a port management system. Predicting a ship's expected time of arrival (ETA) is a vital step in the development of a smart port system. This study aims to develop a data-driven model to estimate the ETA of incoming ships to port Klang of Malaysia based on past voyage data. An artificial neural network (ANN) based model to predict ETA has been proposed in the study which uses the remaining distance following the trajectory, the instantaneous speed and heading of the ship as input parameters. The proposed model achieves a Mean Absolute Percentage Error (MAPE) value of 36.99% with a Mean Absolute Error (MAE) value of 4603.1367 seconds and a Root Mean Square Error (RMSE) value of 14029.6972 seconds. The model's coefficient of determination was calculated to be 78.67% indicating a satisfactory fit to the dataset. The trajectory has been predicted using a Kalman filter and this predicted trajectory has been used as the input to the neural network model in order to provide a holistic solution to the problem.

Chapter 1

Introduction

1.1 The Estimated Time of Arrival

The Estimated Time of Arrival (ETA) is a common phrase in public transportation, used to describe the time when a transport is expected to arrive at a certain destination. It is often paired with ETD or the Estimated Time of Departure, which expresses the expected start time of a transport's journey.

ETA prediction is an essential task of transportation and logistics companies. In transportation, accurate prediction of bus arrival times lets passengers know if their ride is running late. They may use this information to choose another bus that aligns with their schedule. For airports and railways, advance notices of delays can help decrease losses due to customer dissatisfaction and hampered corporate image. In supply chain and delivery management, accurate arrival time predictions are literally a necessity. Accurate delay predictions enable companies to take pre-emptive measures to manage production downtime and stock shortages. It allows delivery companies to manage their schedules so that dispatchers deliver on time, which helps ensure customer satisfaction. In shipping, arrival time predictions enable port management to draw holistic schedules by taking early arrivals, delays, berth availability and available staff into account. Our research focuses on this field of ETA prediction.

The shipping industry plays a crucial role in trade and commerce as around 80% of the global trade logistics is carried out by ships [1]. So proper management of port resources is essential for the economy of a country. Knowing ship arrival times, ports can minimize vessel turnaround times and predict when vessels will arrive at berths. If berth mooring time is known, operators can optimize cargo operations such as loading and unloading goods, and better allocate resources to maximize work efficiency at port terminals. Mariners may use this information to adjust their vessel's route so that they arrive on time. Cargo owners and port logistics use this information to plan product storage and delivery.

Unexpected ship arrivals cause congestion in the port area. Congestion is the situation where vessels arriving at a port already at max capacity have to queue outside the port area for a spot to load/unload. For example, suppose that a container ship is scheduled to arrive on Sunday at Chittagong port ended up arriving on Monday due to bad weather. The port was not informed of this delay so it did not have any berths available at that time. This forced the container ship to anchor outside the port area for a few hours. Just one ship anchoring outside the port is not that big of an issue but when multiple ships do this, it causes some major problems.

1.2 Port Congestion and Schedule Unreliability

With the increase in container sizes and their carrying capacity, port congestion is becoming a bigger problem by the day, and it currently accounts for 93.6% of the port delays [2]. A study in [3] found that only 52% of the vessels dispatched for liner services arrived at the planned schedule. This study also elaborates on how schedule unreliability affects several actors throughout the supply chain. Delays in ports add to the total duration of a shipping line's round-trip time. This affects profits by costing additional daily ship fixed costs. To make up for lost time, the ship will have to sail at full speed, which will further decrease profits by increasing the ship's operational cost. Unreliability of arrival time complicates the process of berth planning and yard planning for terminal operators, as it is difficult to allocate port space and resources with large time windows to adjust delays (need better phrase). Low schedule reliability also affects inland transport operators, causing delays in their operations and reducing their productivity level. It increases logistical costs of the customers as they have to invest in higher inventory levels in order to avoid disruptions to production processes. Unproductive vessel time is also undesirable for the environment. A study in [4] shows that vessel emission concentration of CO and CO2 peak when vessels are anchored.

It is very easy for ship schedules to be disrupted because of the various uncertainties at sea. Accurate predictions of the ETAs of shipping lines help port operators handle unexpected arrivals and thus ensures efficient port management and optimal use of port facilities. Our research focuses on developing an efficient model for predicting the ETA of ships using ANNs (Artificial Neural Networks) and Kalman filters.

1.3 Existing ETA prediction methods and problems

The most rudimentary form of ETA calculation, still employed by some shipping lines, is done by dividing the distance to port by the ship's speed over ground (SOG). This form of estimation is often erroneous as it does not take external factors into account. So recent literatures focusing on ETA prediction propose methodologies based on pathfinding and Machine Learning (ML) algorithms trained on historic Automatic Identification System (AIS) data. While they do produce acceptable results, most of these data-driven methodologies are not suitable for continuous real-time predictions. Therefore, this study aims to present an accurate data-driven methodology that will be able to predict ship arrival times in real-time using a Kalman filter and an Artificial neural network (ANN), as shown in Fig. 1.1. The model implements a Kalman filter that predicts the ship's trajectory. From the trajectory, the model derives the remaining distance to be covered. This feature is used as one of the inputs of the ANN part of the model, which then predicts the ETA. Port planners may use the predicted ETA to generate cost-effective port schedules. Such an intricate system will assist smart ports in facing the challenges of the ongoing 4^{th} Industrial Revolution.

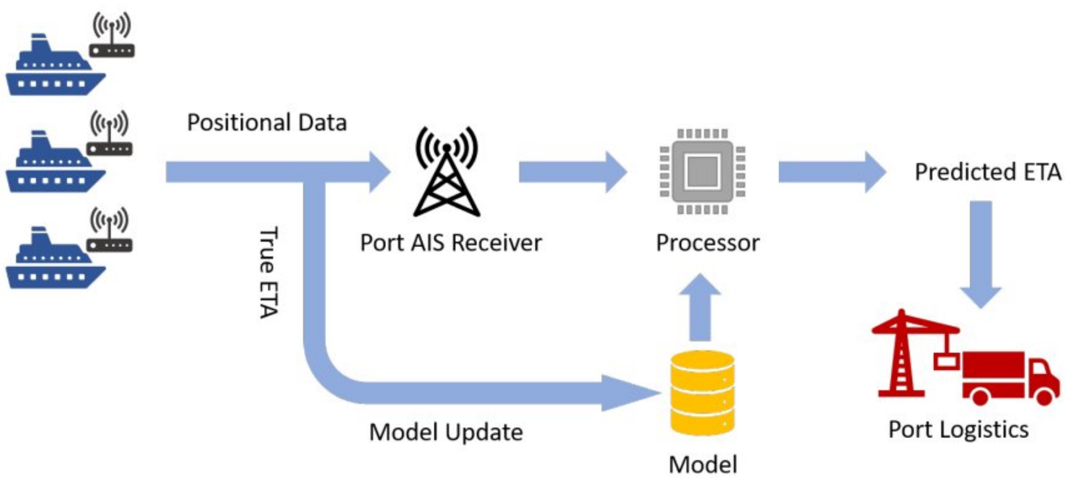

Figure 1.1: Overview of the proposed system

Chapter 2

Literature Review

In this chapter we will discuss some of the research related to ETA prediction. Methodologies for ETA and route prediction using kalman filters and ML techniques will be explored.

2.1 Kalman Filters in ETA Prediction

Kalman filters are particularly popular in bus arrival time predictions. In [5], Kalman filters have been used to predict bus arrival times through a data-driven methodology that exploits the special and temporal correlations in travels times using real world data taken from Global Positioning System (GPS) fixtures. In [6], a dynamic model is used to predict bus travel time on roads with multiple bus routes. The study used an SVM (Support Vector Machine) trained on historical data to estimate the baseline bus travel time on roads with multiple routes. Then a Kalman-filter based algorithm was used to adjust the prediction of the first component using the latest travel time information. [7] uses a similar process to predict bus arrival times, but with an ANN instead of the SVM used in the previous research.

2.2 Machine Learning in ETA Prediction

As for statistical and ML (Machine Learning) methodologies related to ship ETA prediction, researchers in [8] proposed algorithms that use trajectory mining to determine the ship's route to predict the ETA from a point on the map to a port. In [9], a Markov decision process (MDP) based reinforcement learning (RL) framework is used for optimal pathfinding and the Metropolis-Hastings algorithm to estimate the SOG.

The ship's ETA was calculated by dividing the distance of the predicted trajectory by the estimated SOG. Another study proposed a path-finder that exploits historical ship tracking data in [10]. Then they extracted the ship's SOG from its AIS message to predict ETA. In [11], authors map the entire Mediterranean Sea in a spatial grind. A seq2seq model predicts the ship trajectory in the form of a sequence of cells. Then the average acceleration of ships moving from cell to cell is used to predict the arrival time.

These studies show that both Kalman filtering and ML techniques have strong foundations in travel time prediction. But in terms of performance, Kalman filters lag behind. The same is observed in [6] and [7], where authors chose ML techniques to predict the baseline ETA, using Kalman filtering techniques for error adjustment instead.

2.3 Kalman Filters in Position and Route Estimation

Kalman filters are popular for their applications in route and position prediction. In [12], Kalman filtering techniques have been used to predict the position of vehicles in VANETs (vehicular ad-hoc network). In [13], the trajectory of an ocean vessel is predicted by using an extend Kalman filter algorithm to estimate the system states of position, velocity and acceleration of the vessel. [14] uses multiple Kalman filters to model the different behaviors of a moving vehicle. The paper separates vehicle behavior into four possible states – constant location, constant velocity, constant acceleration and constant jerk. A comprehensive Kalman filter to correctly identify the state of the vehicle, then the corresponding Kalman filter model is used to estimate the vehicle's trajectory. [15] uses a Kalman filter to model a system to be used by robots to predict the future position and orientation of moving obstacles from obstacle data collected via sensors. Kalman filters do not require to be trained using past data. Even with state variables having zero initial values, Kalman filters are capable of learning system behavior from small samples of data. Thus, Kalman filters can be reliably used to predict the future position and trajectory of a moving object.

The methodologies that have been proposed are multi-step processes that are too slow for real-time prediction. This research proposes a hybrid ANN and Kalman filter model that is fast enough to predict ETA of ships in real time.

Chapter 3

Background

This chapter will go over the details of the theories related to our research: AIS, Kalman filters and ANNs.

3.1 Automatic Identification System (AIS)

Vessel Traffic Services (VTS) tracks ships using transceivers and the Automatic Identification System (AIS). It enables ships to monitor adjacent water traffic. As a result, a separate AIS transceiver is required. Because they just need to monitor local traffic and do not need to send their own position, port authorities and other shore-based facilities sometimes simply have receivers. All local marine activity equipped with an AIS transceiver onboard may be viewed this way quite reliably within 20–30 nautical miles, depending on the range of the transceiver [16].

All vessels above 300GT on international journeys must have a class A type AIS transmitter, according to the 2002 IMO SOLAS agreement [16]. This was the first AIS regulation, which impacted over 100,000 vessels. The class B type AIS transceiver specification was released by the AIS standards committee in 2006. It was intended to be a less expensive AIS gadget. In the same year, Class B transceivers were commercially accessible, prompting several governments to enact laws. This reduced the cost of installing AIS systems on boats of various sizes.

The AIS technical standard committees have been expanding the AIS standard and product types since 2006 to enable a broad range of applications. Simultaneously, governments and authorities have started a number of programs to equip different types of boats with AIS devices in order to improve safety and security.

Although the primary goal of AIS was to prevent collisions, it has lately been expanded to include a wide range of additional uses. Collision avoidance, monitoring and control of fishing fleet, navigational aid, maritime security, accident investigation, search and rescue, ocean current estimates, infrastructure protection, fleet and cargo tracking, statistics and economics, and so on are all examples of where AIS is currently used.

According to [16], some data are sent by AIS transceivers every 2 to 10 seconds when the vessel is in motion and every 3 minutes when the vessel is anchored. The transmissions during motion vary with the speed of the vessel. Some of these information are:

- Vessel Maritime Mobile Service Identity (MMSI) number: It is a unique nine-digit identification number assigned to all vessels.
- Navigation status: Indicates the current condition of a vessel. Some of the key-words under this field are, "at anchor", "underway using engine(s)", "not under command", etc.
- Rate of turn: This field provides infomation about the angular velocity of a ship. A turn can be right or left, with possible values from 0 to 720 degrees per minute
- Speed over ground: Provides information about the linear component of a vessel's motion. Minimum resolution of the field is 0.1-knot (0.19 km/h). And it typically ranges from 0 to 102 knots (189 km/h)
- Positional resolution: Provides the current geographical co-ordinates of a vessel in longitude and latitude. The minimum resolution for both parameters are typically 0.0001 arcminutes.
- Course over ground: Provides angular information of the vessel relative to true north. Minimum resolution is limited to 0.1°
- UTC time

Along with the above mentioned information, some extra data are broadcast every 6 minutes. Some example of these data are as follows:

- International Maritime Organization (IMO) ship identification number: It is a seven-digit number used to uniquely identify a vessel. Even if the ship's registration is transferred to another nation, the IMO number stays the same.
- Radio call sign: The country of registry assigns the vessel an international radio call sign. The maximum length of this field is 7 characters.
- Name of the vessel: This field can have a maximum of 20 characters.

- Type of vessel: Transmitted as a word. Some common examples are - 'Cargo', 'Container', 'Oil Tanker', etc.

- Draught of ship: This is the vertical distance between waterline to the vessel's bottom and it generally depends on the amount of weight carried by the vessel. The typical measurement ranges from 0.1 to 25.5 meters.

- Destination: Destination is expressed by a maximum of 20 characters

- High precision time request: This is an optional information. A ship can request that some other vessel supply a high-precision UTC time and date stamp, and only if this request is acknowledged is this information sent.

For information exchange, AIS equipment uses NMEA 0183 phrases. For AIS data, the NMEA 0183 standard includes two major phrases [17]. The !AIVDM phrase is used when the received data originates from other vessels and the !AIVDO phrase is used while transmitting the vessel's own information.

The AIS data encoded messages are also termed as AIVDM or AIVDO sentences. A sample of AIS encoded messages is shown in Fig. 3.1. Each coded sentences of the simplest form have seven commas separated placeholders. More details on AIS encoding can be found in [18]

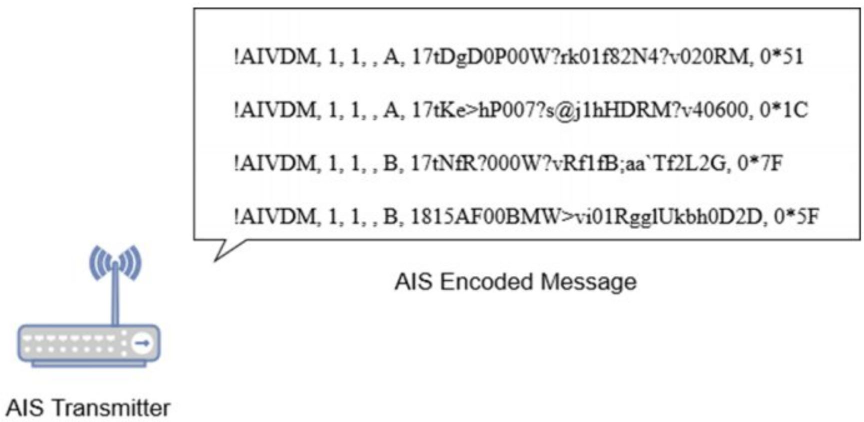

Figure 3.1: Examples of encoded AIS message

3.2 Kalman Filter

The ship's trajectory was predicted using the Kalman filter approach. The method is iterative. It may operate in real time, with just the current input measurements, the previously established state, and the uncertainty matrix as inputs; no prior knowledge is required. A block diagram of a typical Kalman filter is shown in Fig. 3.2.

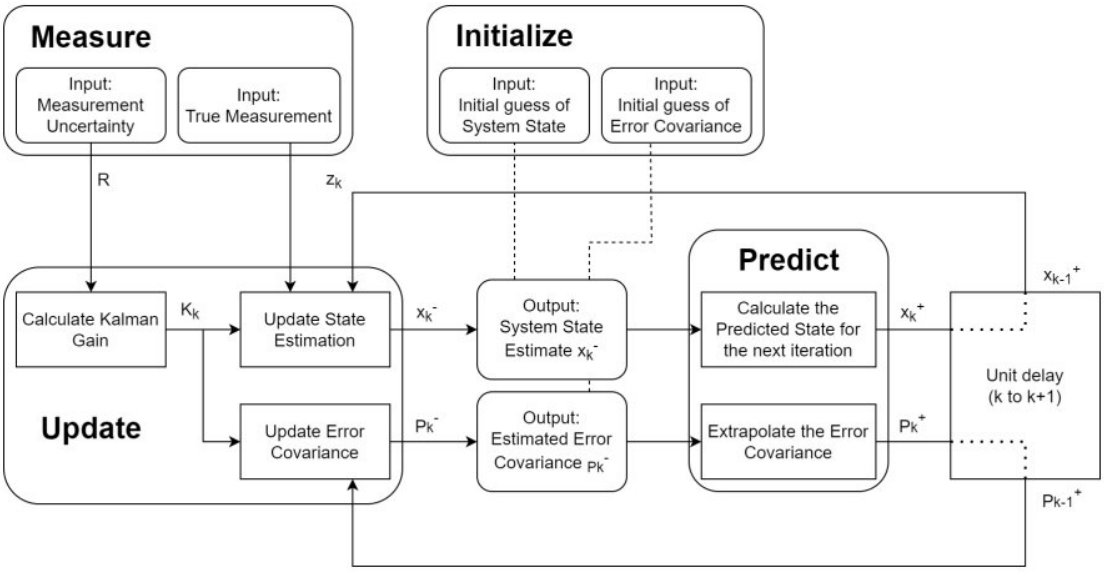

Figure 3.2: A typical Kalman filter

The Kalman filtering algorithm uses a series of measurements taken over time, including statistical noise and other inaccuracies, to produce estimates of unknown variables that are more accurate than those based on a single measurement alone, by estimating a joint probability distribution over the variables for each timeframe. The filter is named after Rudolf E. Kálmán, who came up with the concept in 1960 [19].

Originally the Kalman filter was proposed as a recursive solution to the discrete data linear filtering problem [19]. It gained fame because of how simple it was to utilize with digital computers. Because understanding the theory underlying the filter's construction was not required in order to apply it, it became a hotbed of study and application. Nowadays, Kalman filtering has a broad variety of technical applications, particularly in autonomous or aided navigation.

Kalman filters are extensively used in airplanes, spacecraft, and ships for guidance, navigation, and control [20–22]. Furthermore, Kalman filtering is a widely used time series analysis approach with applications in signal processing and econometrics [23]. Kalman filtering is one of the key themes in robotic motion planning and control, and it may be used for trajectory optimization, as discussed in Section 2.3.

There are two steps to the algorithm. For the prediction phase, the Kalman filter produces estimates of the current state variables in state space format, together with their uncertainty. Once the results of the next measurement (necessarily tainted with some error, including random noise) are obtained, these estimates are updated using a weighted average in the following step, with greater weight given to more definite guesses.

In the following equations, the hat operator ˆ denotes an estimate of a variable. The superscript - denotes prior(predicted) and + denotes posterior(updated) estimates from time $k-1$ to k.

3.2.1 Prediction

The following equations are used in the prediction phase of the algorithm:

1. Predicted state estimate,

$$\hat{x}_k^- = F\hat{x}_{k-1}^{+} + Bu_{k-1} \tag{3.1}$$

   Here, x_k is the state vector x at time k, F is the transition state matrix and u_{k_1} is the control vector with B being its control-input matrix. This is the first step in the Kalman filtering algorithm.

2. Predicted error covariance,

$$P_k^- = FP_{K-1}^{+}F^T + Q \tag{3.2}$$

   P is the state error covariance. The error covariance that the filter believes the estimate error has is encoded by this term. The summation of the process noise covariance matrix, Q increases the error covariance at this stage, indicating that the filter is now more doubtful about the state estimate $\hat{x}$.

3.2.2 Update

The steps of the updating process comprise the following equations:

1. Measurement residual,
$$\widetilde{y_k} = z_k - Hx_k^- \tag{3.3}$$
$\widetilde{y_k}$, the measurement residual, is calculated first as the difference between z_k, the true measurement, and the estimated measurement, Hx_k^-. H denotes the measurement matrix that is multiplied to the predicted state.

2. Kalman gain,
$$K_k = P_k^- H^T \left(R + HP_k^- H^T\right)^{-1} \tag{3.4}$$
Here, K_k is the Kalman gain at time K. R is the measurement noise covariance matrix, generally calculated during the calibration of measuring equipment.

3. State estimate,
$$\hat{x}_k^+ = \hat{x}_k^- + K_k\widetilde{y} \tag{3.5}$$
Then we multiply the residual $\widetilde{y}$ by the gain K_k to provide the correction $K_k\widetilde{y}$ to the predicted estimate $\hat{x}_k^-$. Then $\hat{x}_k^+$ is the updated estimate.

4. Error covariance,
$$P_k^+ = (I - K_k H)\, P_k^- \tag{3.6}$$
The final step is to update the error covariance. The updated error covariance P_k^+ in this step is smaller than the predicted error covariance P_k^- which signifies that the filter is now more certain of the state estimate.

3.2.3 Initialization of variables

To implement the Kalman filter, we need to initialize the values of the state estimate $\hat{x}_0^+$ and error covariance matrix P_0^+. For quicker convergence, a large P_0^+ may be chosen. These guesses play important roles in obtaining the desired performance and thus should be chosen carefully.

3.3 Artificial Neural Networks (ANN)

An ANN is modeled based on the biological neuron. Like the biological neural network, artificial neural network is an interconnection of nodes that have a certain weight and a bias which are trained through backpropagation. In an ANN structure there are three types of layers – the input layer, the output layer and one or multiple hidden layers. A typical ANN structure can be visualized in Fig. 3.3 and a node structure can be visualized in Fig. 3.4.

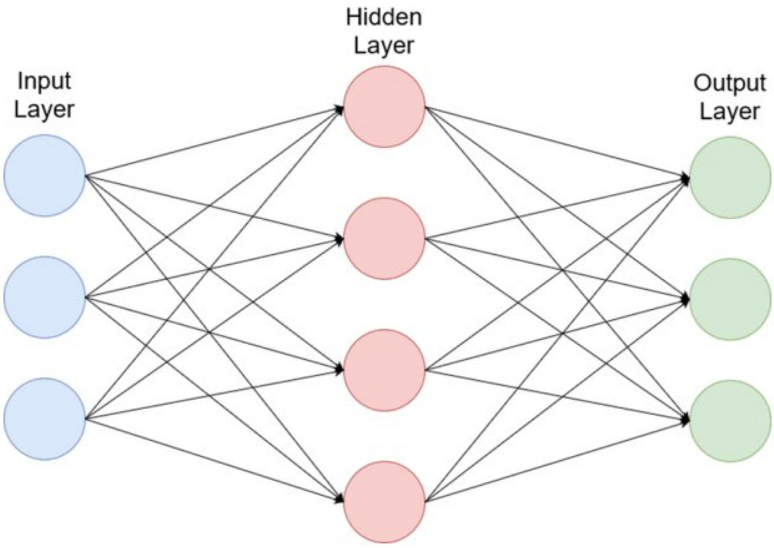

Figure 3.3: Typical ANN configuration

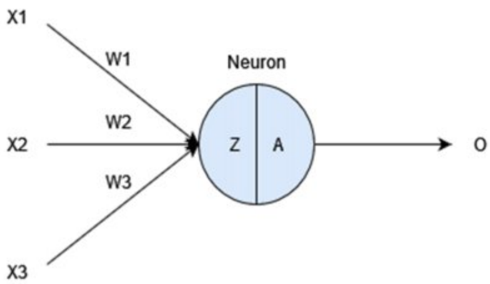

Figure 3.4: Typical Neuron Structure of an ANN

The forward function of an ANN is defined by:

$$O = g(wX + b), \tag{3.7}$$

Where, w is the weight matrix and b is the bias matrix. X is the input matrix and $g(x)$ represents the activation function. There are different types of activation functions. Some of the most common activation functions are:

Sigmoid activation function,

$$g(x) = \frac{1}{(1 + e^{-x})} \tag{3.8}$$

Rectified Linear Unit (ReLU) activation function,

$$g(x) = max(0, x) \tag{3.9}$$

and the Hyperbolic tangent (tanh) activation function,

$$g(x) = \frac{(e^{2x} - 1)}{(e^{2x} + 1)} \tag{3.10}$$

For an ANN backpropagation is mainly done using different forms of gradient descent algorithms. In recent literatures, different metaheuristic algorithms are being used to get the maximum optimization. In this paper we have limited ourselves to regular optimization algorithms. The most common gradient descent algorithm is given by the function,

$$X = X - learningrate * \frac{\mathrm{d}f(X)}{\mathrm{d}x} \tag{3.11}$$

Where X is the weight value to be optimized and $f(X)$ is the output value for the weight of X and learning rate is a hyperparameter set by the user.

Chapter 4

Methodology

In the context of Industry 4.0, the port has to undergo a significant change to accompany the global economic growth. In our ports, ship schedulers are continuously balancing a multitude of variables to prepare the vessels' loading and unloading schedule. Congestion and unreliability of vessel arrival times further complicates the scheduling process. The schedulers have to adjust loading and unloading windows at port terminals to account for the delays caused by congestion and unexpected arrivals. This is a trial and error method that makes it difficult for the port authority to make optimal use of port resources.

The aim of this research is to help port operators draw holistic schedules by removing the uncertainties in arrival times via ETA prediction. A ship schedule is an essential part of port management that controls the final port rotation. If the arrival time of vessels are knows, the information may be used to generate optimized port schedules that minimize terminal idle time and maximize the port's profits. An efficient schedule will also allow ports to avoid congestion.

We initially started with a standalone ANN model with the location, speed and course of a vessel as input features. The model failed to deliver any acceptable result, having a MAPE over 70%. On further data analysis we found that adding the remaining distance as an input drastically raises the accuracy of the model. The remaining distance attribute we used then was pre-processed from the collected data. It goes without saying that this model was unusable in real time. So we needed a system that could predict the distance to be covered by a ship in real time. Through extensive literature review, we chose a Kalman filter for this task. The prediction model we are proposing in our thesis is a hybrid Kalman-ANN system. The ANN part of the model has to be pre-trained using historic AIS data of ships arriving to and departing from the concerned port. For our thesis, we worked with the port Klang of Malaysia. The ANN was trained using data containing AIS messages received at port Klang during the first month of 2019.

The dataset we used contained encoded AIS messages. These were then decoded using the pyais Python library. The decoded data was sorted and processed to retain only useful information related to the voyage of incoming ships. A Kalman filter is applied to predict the ship's trajectory from the extracted data. From the predicted trajectory, the remaining distance on the ship's route is calculated. This is done for every data point. The data is then split into training, testing and validation sets. The training set is used to train the ANN while the testing and validation sets are used for model evaluation. The overall workflow is depicted in Fig. 4.1 and discussed in detail in the following chapters.

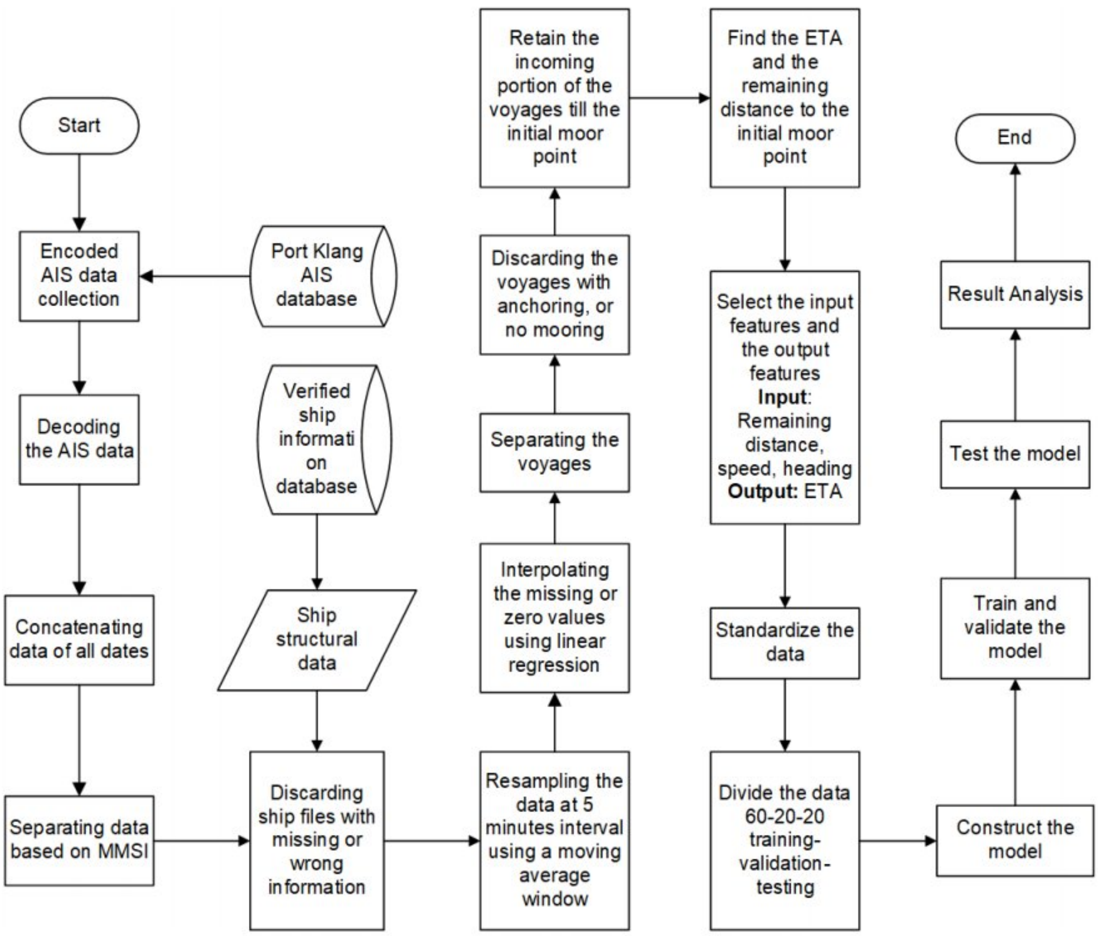

Figure 4.1: Workflow for the prediction of ship arrival times

In real time, all required data will be decoded from live AIS transmissions received by AIS receivers installed at the concerned port. A Kalman filter will estimate the remaining distance in the ship's trajectory from the received data. This data along with the attributes speed and course of the ship will be inputs to the ANN system that will predict the ship's ETA. A simple schematic of the proposed system is shown in Fig. 4.2.

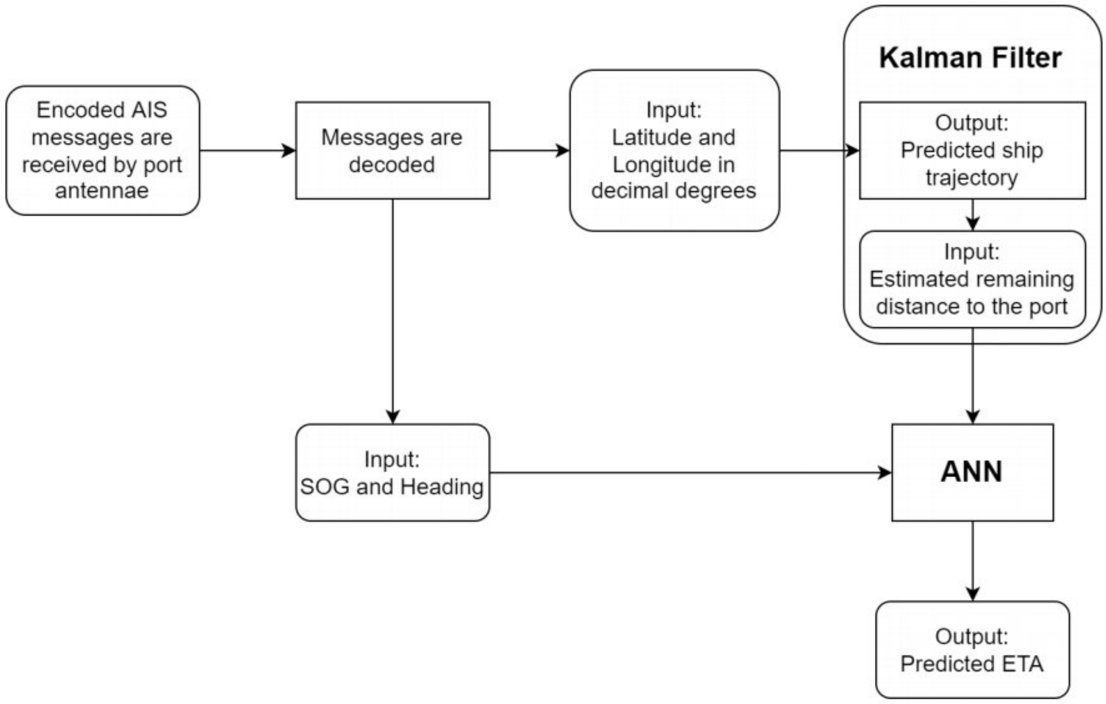

Figure 4.2: A schematic overview of the proposed system

The proposed system will help port operators plan efficient port schedules. This will enable shipping companies to improve their shipping operation, manage facilities optimally, reduce idle time and maximize profits by allowing more ships.

Chapter 5

Data Collection and Preprocessing

In this chapter we will discuss all steps related to data collection ad preprocessing in detail.

5.1 Data Collection

The data was collected from Malaysia's Port Klang's authority. The dataset was comprised of all the voyages to and from the port for one month starting from 1^{st} January, 2019 to 1^{st} February, 2019. The data was recorded by the port's AIS transceiver which covered a radial distance of 100 km from the port.

The dataset was obtained in CSV format. Unfortunately, the received dataset wasn't exported properly which corrupted the AIS codes. An example of received code:

$$!AIVDM|1|1||A|18KDFp1Oh6W?r5J1hDRhGAV60H1f|0|3D|18KDFp1Oh6 W?r5J1hDRhGAV60H1f|KAZAK|1|0|565516000|565|1|127|0.6|1|101.342200| 3.067112|9.3|51|03|0|0|0|0|0|6|110$$

Whereas a properly encoded AIS code looks like:

$$!AIVDM,2,1,3,B,55P5TL01VIaAL@7WKO@mBplU@ < PDhh000000001S ;AJ :: 4A80?4i@E53,0*3E$$

It can be clearly observed that multiple delimiters have been changed to a vertical slash '|'. This caused problem when the data was tried to be decoded. A trial-and-error method was adopted to recover the data by force changing the delimiters and trying to decode the updated codes. After a few trials most of the codes could be deciphered. There were two forms of AIVDM sentences in the dataset and this method worked for one type only which accounted for 80% of the data. The rest of the data was lost and they were not considered for the study.

The decoded data was validated by plotting them on the map using GeoPandas library in Python. Only the geographically consistent records were accepted and the rest were discarded. A further consistency check was done by gathering the physical information of each vessel from verified maritime tracking websites [24] and [25] and cross checking the corresponding SOG and COG reading of the vessels. The inconsistent vessels were dropped from the study.

5.2 Data Preprocessing

At first the data was separated into ships by their MMSI numbers. The recovered data had a lot of outliers and noisy readings for example, erroneous longitude and latitude reading, impossible SOG and COG values, trajectory going over land area. These apparent outliers were removed firstly. Then the structural data obtained from the verified website was used to further filter the readings based on SOG and COG values. It was observed that for some of the vessels, there were only a few readings. These ship data were deemed unreliable and they were not considered for the study.

Figure 5.1: Before Processing

Before processing, the dataset contained trajectories as clusters of data points sampled at non-uniform intervals, with some parts of the trajectory not recorded at all as shown in Fig. 5.1. The dataset at this point was not sampled uniformly as AIS transceivers don't transmit data maintaining a fixed delay. But for application in a model, uniform data would perform better than non-uniformly sampled data. As such, we resampled the data using a moving average window. This smoothened the trajectories and helped eliminating small errors.

Figure 5.2: After Processing

Figure 5.3: After Interpolation

Now the problem that we have to address is that there are certain missing points in the trajectory as seen in Fig. 5.2. In order to impute these missing points, we apply a simple linear interpolation which gives us the trajectory in Fig. 5.3. The ship data available at this point of preprocessing has been provided in the Table 5.1 and a pie chart has been provided in Fig. 5.4.

Table 5.1: Ship distribution by type

Type	Count
Cargo	87
Carrier	61
Container	355
Fishing	18
Passenger	11
Tanker	94
Tug	46
Other	20
Unspecified	3
Total	**695**

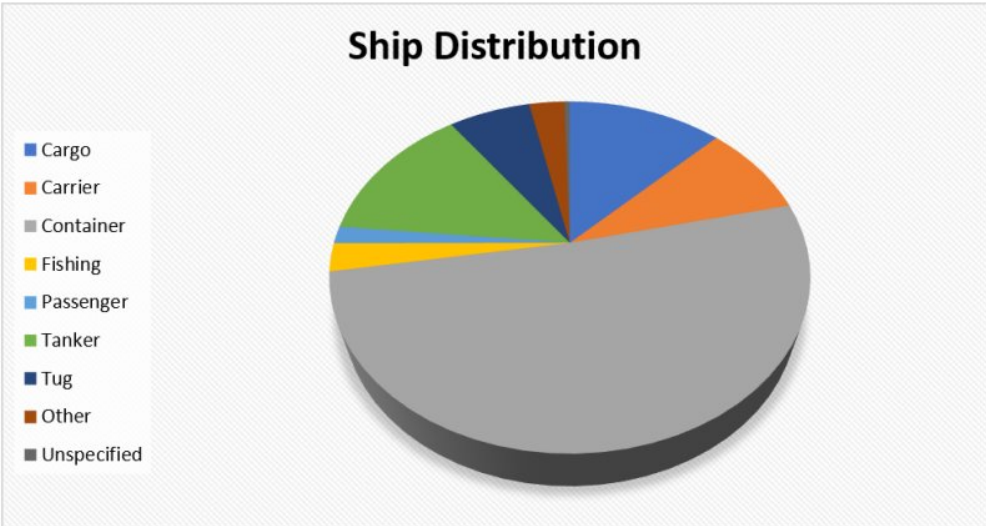

Figure 5.4: Ship Distribution Graph

The final step of preprocessing involved dividing the ships into voyages. Since we are considering data for one month, some ships have had multiple voyages in and out the port. For this specific study we have considered the ships that haven't anchored at any point of their voyages as anchor time falls under the study of port logistics. Furthermore, we have only considered the international vessels at this point of preprocessing. Ultimately, we extracted 645 voyages from a total of 620 ships.

Chapter 6

Model Creation and Training

6.1 Estimating the remaining distance

The first part of the proposed model is a Kalman filter that predicts the trajectory of ships. The filter was developed using Python programming with generic libraries. The latitude and longitude of the ship's geographic position are considered as system states. These values are taken from transmitted AIS messages. The filter is tested using data sampled at 1, 5, 10 and 15 minute intervals. In each iteration, the states latitude and longitude of the next sampling point are predicted using the system equations no. 3.1 to 3.6. Using the predicted point, the filter predicts the states at the following sampling point and so on. In this manner, the entire trajectory is predicted. Following the points in the predicted trajectory, the remaining distance to the port is estimated. To better visualize the filter output, a sample predicted trajectory is plotted in Fig. 6.1. White points represent the predicted positions while blue points are the actual positions.

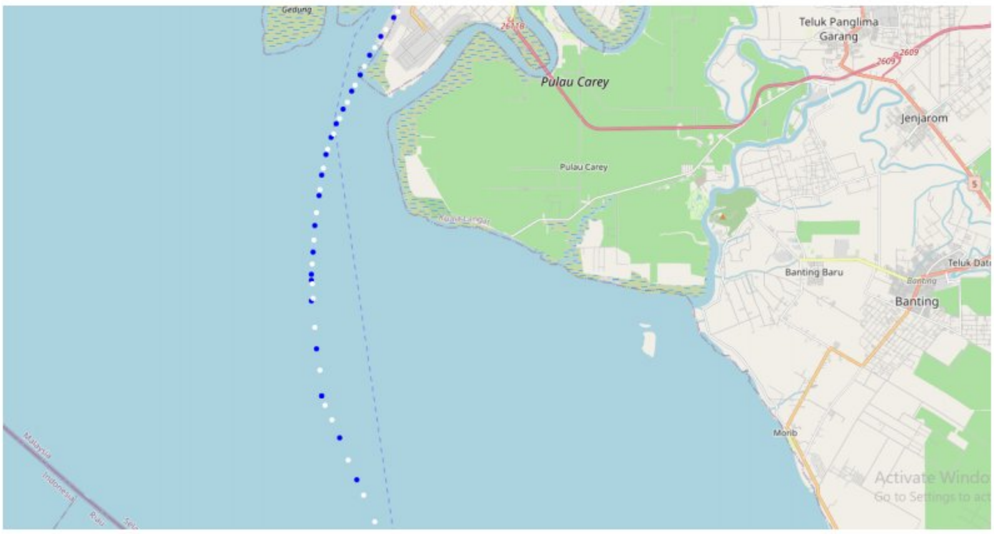

Figure 6.1: A plot showing a predicted trajectory

6.1.1 Estimation results

In the following Tables no. 6.1 to no. 6.4, the mean error in the estimation is tabulated in the form of mean distance between the predicted ship positions and the actual ship positions in data:

1. Sampled at 15 minutes

Table 6.1: Sampled at 15 minutes

Minutes ahead of prediction	Mean Distance (meters)
15	456.7403
30	1220.115
45	1700.447
60	2172.775
75	2652.011
90	3073.621

2. Sampled at 10 minutes

Table 6.2: Sampled at 10 minutes

Minutes ahead of prediction	Mean Distance (meters)
10	306.509
20	949.6295
30	1229.683
40	1562.841
50	1861.765
60	2187.302

3. Sampled at 5 minutes

Table 6.3: Sampled at 5 minutes

Minutes ahead of prediction	Mean Distance (meters)
15	747.0585
25	1178.977
35	1579.848
45	1936.38
55	2291.548
60	2487.548

4. Sampled at 1 minute

Table 6.4: Sampled at 1 minute

Minutes ahead of prediction	Mean Distance (meters)
10	605.1668
15	824.6228
20	1122.168
25	1388.096
30	1255.144
35	1843.38

The errors tabulated above are plotted in Fig. 6.2. Here, the performance of the filter measured with data sampled at 10 minute and 15 minute intervals are quite similar with the latter being the better alternative.

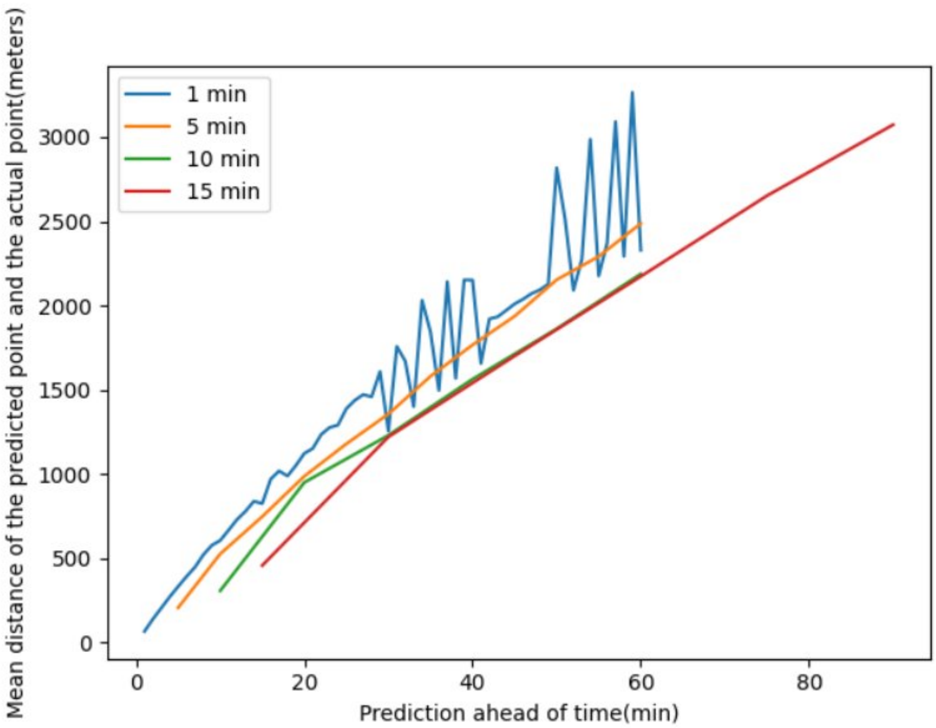

Figure 6.2: Comparison of variations in prediction at different sampling rates

6.2 The ANN model

For the neural network model three parameters are considered as input, the distance remaining, the speed, and the heading value at any given point whereas the ETA is the output. To find the remaining distance to be covered by the vessel a kalman filter is used to predict the path it would follow and from the path the distance to be covered is calculated. Then this along with the heading and speed of the vessel is fed into the neural network to predict the ETA.

All the values of training, validation and test set are standardized. For the model, a sequential ANN is used. The layers and their varying hyper-parameters are listed in Table 6.5.

Table 6.5: ANN layers of the proposed model

Layer Type	**No. of Units**	**Regularizer**
Fully Connected	2048	-
Fully Connected	1024	L2 Regularizer (0.02)
Fully Connected	512	L2 Regularizer (0.02)
Fully Connected	256	L2 Regularizer (0.02)
Fully Connected	128	L2 Regularizer (0.02)
Batch Normalization	-	-
Fully Connected	128	L2 Regularizer (0.02)
Fully Connected	64	L2 Regularizer (0.02)
Fully Connected	64	L2 Regularizer (0.02)
Fully Connected	32	L2 Regularizer (0.02)
Fully Connected	16	L2 Regularizer (0.02)
Fully Connected	8	L2 Regularizer (0.02)
Fully Connected	4	L2 Regularizer (0.02)
Fully Connected	2	L2 Regularizer (0.02)
Fully Connected	1	-

For all the fully connected layers, the activation function used was ReLU activation. The loss function was chosen to be based on mean absolute error (MAE). For this model we used the Adam optimization function with a learning rate of 0.001. The early stop callback was used and the model trained for 2187 epochs before stopping. The dataset was divided into 60% training, 20% validation and 20% test set for the problem. To avoid overfitting L2 regularizer has been extensively used in our model. The batch normalization layer has been used to speed up the learning process and combat overfitting of parameters. The early standardization of data also speeds up the model significantly.

Chapter 7

Model Evaluation and Results

7.1 Evaluation Metrics

The model was evaluated on four different parameters, mean absolute error (MAE), mean root squared error (RMSE), mean absolute percentage error (MAPE) and coefficient of determination (R^2 score). The mathematical expressions for these metrics are given as follows:

Mean Absolute Error,

$$MAE = \frac{\sum_{i=1}^{n} |F_i - A_i|}{n}, \tag{7.1}$$

Root Mean Square Error,

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} |F_i - A_i|^2}{n}}, \tag{7.2}$$

Mean Absolute Percentage Error,

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|, \tag{7.3}$$

Coefficient of Determination,

$$R^2 = 1 - \frac{\sum_i (A_t - F_t)^2}{\sum_i (A_t - \hat{A}_t)^2}. \tag{7.4}$$

Where n is the number of samples, F_i is the ith forecasted value and A_i is the ith actual value. For R^2 the term $\hat{A}_i$ stands for the mean of the ith sample. The lower the value of MAE and RMSE, the better the model. MAPE restricts the error within 100% and the lower the value is the better the model is. Finally, the coefficient of determination or R^2 score has a range of 0 to 1. The higher the value, the better the model performs on the given task.

7.2 Model Performance

During the training, the model didn't overfit as observed from the training and validation loss curves provided in Fig. 7.1. The validation loss curve has closely followed the training loss curve before the training was terminated. The performance results of the model are tabulated in Table. 7.1

Table 7.1: Error metrics

	Train Set (60%)	Validation Set (20%)	Test Set (20%)
MAE	4082.12 seconds	4606.26 seconds	4603.14 seconds
RMSE	13014.96 seconds	13391.15 seconds	14029.70 seconds
MAPE	31.56%	34.61%	36.99%
R^2	-	-	0.7867

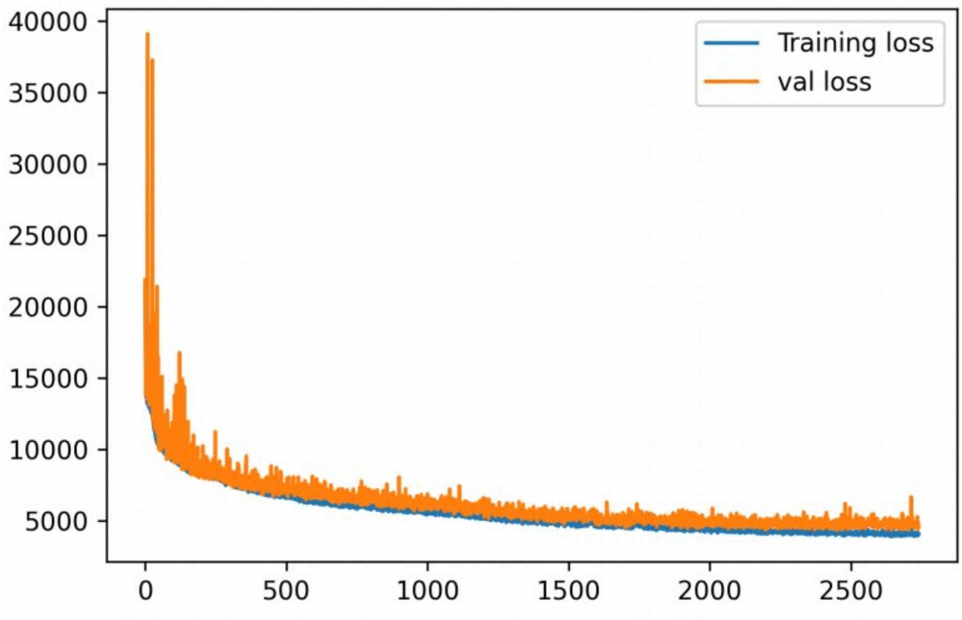

Figure 7.1: Training and validation loss curve

Chapter 8

Result Discussion and Conclusion

This study has proposed a data driven ANN based model to predict the ETA of ship using information from its AIS transmissions. The model uses only three parameters - distance remaining to the destination following the trajectory, current speed over ground and the current heading value to predict the output. Predicting the trajectory has been treated as a separate problem and has been addressed before applying the ANN model. As a solution a Kalman filter based trajectory prediction system has been developed.

The ETA prediction model and the trajectory prediction model both have been trained using real data obtained from Malaysia's port of Klang authority recorded over a month. The real challenge was decoding the encoded data. The data had severe encoding corruption. The AIS messages were force decoded by a trial and error method. The data was validated by plotting them on the map and sifting through the inconsistencies. A rigorous preprocessing step was performed including re-sampling and imputations of missing data points before the data was used in the model.

The final result has shown very good performance by the model, achieving an MAE of 4603.1367 seconds with an MAPE of 36.99% and an RMSE value of 14029.6972 seconds. The coefficient of determination obtained for the complete model was 78.67% which indicates a good fit to the dataset. The fact that the model works with only three parameters gives it an edge over other prediction methods for its fast computation and its ability to be implemented in real-time operation. This opens up a future scope of work where the algorithm can be implemented on hardware using embedded technologies for real-time use. Further development on the ETA prediction can be done by incorporating a port logistics model with the proposed model which can extend the domain of application manifold.

REFERENCES

[1] United Nations Conference on Trade and Development, “Review of Maritime Transport 2021.” unctad.org. https://unctad.org/webflyer/review-maritime-transport-2021 (accessed May 6, 2022).

[2] T. Notteboom, “The time factor in liner shipping services,” *Maritime Economics and Logistics*, vol. 8, pp. 19–39, 03 2006.

[3] B. Vernimmen, W. Dullaert, and S. Engelen, “Schedule unreliability in liner shipping: Origins and consequences for the hinterland supply chain,” *Maritime Economics and Logistics*, vol. 9, pp. 193–213, 09 2007.

[4] C. Bai, Y. Li, B. Liu, Z. Zhang, and P. Wu, “Gaseous emissions from a seagoing ship under different operating conditions in the coastal region of china,” *Atmosphere*, vol. 11, no. 3, 2020. [Online]. Available: https://www.mdpi.com/2073-4433/11/3/305

[5] A. Achar, D. Bharathi, B. A. Kumar, and L. Vanajakshi, “Bus arrival time prediction: A spatial kalman filter approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1298–1307, 2020.

[6] C. Bai, Z.-R. Peng, Q.-C. Lu, and D. J. Sun, “Dynamic bus travel time prediction models on road with multiple bus routes,” *Computational intelligence and neuroscience*, vol. 2015, p. 432389, 07 2015.

[7] M. Zorkany, M. Zaki, I. Ashour, and B. Hisham, “Online bus arrival time prediction using hybrid neural network and kalman filter techniques,” *International Journal of Modern Engineering Research (IJMER)*, vol. 3, 01 2013.

[8] S. Mao, E. Tu, G. Zhang, L. Rachmawati, E. Rajabally, and G.-B. Huang, “An automatic identification system (ais) database for maritime trajectory prediction and data mining,” 07 2016.

[9] K. Park, S. Sunghyun, and H. Bae, “Vessel estimated time of arrival prediction system based on a path-finding algorithm,” *Maritime Transport Research*, vol. 2, p. 100012, 01 2021.

[10] A. Alessandrini, F. Mazzarella, and M. Vespe, “Estimated time of arrival using historical vessel tracking data,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 7–15, 2019.

[11] D.-D. Nguyen, C. Van, and M. I. Ali, “Vessel destination and arrival time prediction with sequence-to-sequence models over spatial grid,” 06 2018, pp. 217–220.

[12] H. Feng, C. Liu, Y. Shu, and O. Yang, “Location prediction of vehicles in vanets using a kalman filter,” *Wireless Personal Communications*, vol. 80, pp. 543–559, 08 2014.

[13] L. Perera and C. Guedes Soares, “Ocean vessel trajectory estimation and prediction based on extended kalman filter,” 11 2010.

[14] C. Barrios and Y. Motai, “Improving estimation of vehicle’s trajectory using the latest global positioning system with kalman filtering,” *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 12, pp. 3747–3755, 2011.

[15] A. Elnagar, “Prediction of moving objects in dynamic environments using kalman filters,” in *Proceedings 2001 IEEE International Symposium on Computational Intelligence in Robotics and Automation (Cat. No.01EX515)*, 2001, pp. 414–419.

[16] International Maritime Organization, "Guidelines for the onboard operational use of shipborne Automatic Identification Systems (AIS)," *International Maritime Organization*, Resolution A.917(22), 2002. [Online]. Available: https://wwwcdn.imo.org. (accessed May 5, 2022).

[17] *NMEA 0183 Interface Standard*, version 4.11, National Marine Electronics Association, USA, 2018.

[18] E. S. Raymond, "AIVDM/AIVDO protocol decoding." gpsd.gitlab.io. https://gpsd.gitlab.io/gpsd/AIVDM.html (accessed May 5, 2022).

[19] R. E. Kálmán, “A new approach to linear filtering and prediction problems" transaction of the asme journal of basic,” 1960.

[20] P. Zarchan and H. Musoff, *Fundamentals of Kalman Filtering: A Practical Approach*, ser. Progress in astronautics and aeronautics. American Institute of Aeronautics and Astronautics, Incorporated, 2000. [Online]. Available: https://books.google.com.bd/books?id=AQxRAAAAMAAJ

[21] J. Balchen, N. Jenssen, M. Eldar, and S. Saelid, “A dynamic positioning system based on kalman filtering and optimal control,” *Modeling, Identification and Control*, vol. 1, 07 1980.

[22] P. Fung and M. Grimble, “Dynamic ship positioning using a self-tuning kalman filter,” *IEEE Transactions on Automatic Control*, vol. 28, no. 3, pp. 339–350, 1983.

[23] E. Ghysels and M. Marcellino, *Applied economic forecasting using time series methods*. Oxford University Press.

[24] "Vessel Database." fleetmon.com. https://www.fleetmon.com/vessels/ (accessed May 5, 2022).

[25] "Vessels Database." marinetraffic.com. https://www.marinetraffic.com/en/data/?asset_type=vessels&columns=flag,shipname,imo,mmsi,ship_type,notes (accessed May 5, 2022).

List of Publications

Conference

1. **MR. Rahman**, **E. Haque**, **S. T. Rahman**, K. H. Kabir, and Y. A. Ahmed, "Modelling of an Efficient System for Predicting Ships' Estimated Time of Arrival using Artificial Neural Network", the 2^{nd} International Conference on Information Technology, Amity University, Noida, UP, India, March 3-4, 2022, Paper 288.