UNIVERSITÄT
DUISBURG
ESSEN

Open-Minded

University of Duisburg-Essen

Chair of Transport Systems and Logistics

Modelling and optimization of ship's fuel consumption using Random Forest Regression (RFR)

Modellierung und Optimierung des Treibstoffverbrauchs von Schiffen mittels Random Forest Regression (RFR)

Author:
Hibatul Wafi
3021919

Supervisor:
M. T. Muhammad
Fakhruriza Pradana

Summer semester 2023
25th May 2023

Contents

1 Introduction

The research on a more efficient operation of vessel is a direction that is being actively pursued in the maritime world. Efficient ship operation translates to a noticeable increase in profitability and sustainability. The Fuel Oil Consumption (FOC) is a determining factor in the ship's operating cost, with fuel cost making up to 75% of the total ship operating cost [1, 2]. Different ship parameters affect the ship's fuel consumption and research shows that the reduction of ship speed is the most economical method to reduce fuel consumption. Precise forecasting and modelling of the ship speed will lead to a significant reduction in a ship's operating cost and could possibly extend the longevity of the ship.

In this data-driven thesis, powerful machine learning method will be utilized for modelling and forecasting of the ship's speed. To develop the model. The data used will be from the voyage of a Hammerhus RoPax ship between the port of Koege and Roenne. The modelling and forecasting using Random Forest Regressor will be emphasis of this thesis. The developed model of the ship's speed is then to be validated and where applicable, further optimized to increase its performance. This model is then to be used to predict the ship's fuel consumption.

2 Theoretical Background

This chapter deals with the past and present research in the relevant area which include literature review. This includes the significance of precise modelling of the ship's speed and its subsequent use in forecasting the ship's operation. The theoretical background of Random Forest Regression will be discussed in this chapter

2.1 Random Forest Regression (RFR)

2.2 Ship speed

2.3 Modelling

3 Research Methodology

In this chapter the methodology used to develop the model will be discussed. The discussion on different parameters in the vessel's journey data will be discussed here. This includes the mining and merging of the features. The method used to develop the ship's speed model will be discussed in this chapter. This consists of the parameter used to develop the model. Ultimately, the model is then used to predict the ship's fuel consumption.

3.1 Data Preprocessing

- Two data sources are imported. `AIS_weather_H_ok2_copy.csv` and `AIS_weather_h_rename_copy.csv`. The information from the latter comma delimited file will be used for calculating the ship Speed Through Water (STW). The information required is the true north current direction. Which is obtained from the vector component of the Northward and Southward current.
- This dataframe will be merged with the main dataframe from the file `AIS_weather_H_ok2_copy.csv`.
- Omission of the journey data between Ronne and Sassnitz
- SOG threshold is applied to omit ship mooring and maneuvering to accurately represent the ship's steady state operation [1–4]. This threshold is selected as 5 knots according to [1]
- The AIS data from June is filtered. This data will be used as validation data to check the model's performance.

3.2 Data Analysis

- The features are represented in a histogram plot. For the feature Current speed, anomaly is detected. Certain spike is detected around 0.01–0.03 m/s. Reasons unknown. The data is retained, including the spike, until a definitive answer can be found.
- OPEN QUESTION : What is the necessity of feature standardization / normalization ? Normalization is required for ANN as model training requires the value between 0 and 1. But in case of RFR, there is no such requirement. Through testing, data standardization also does not seem to improve the model's performance.

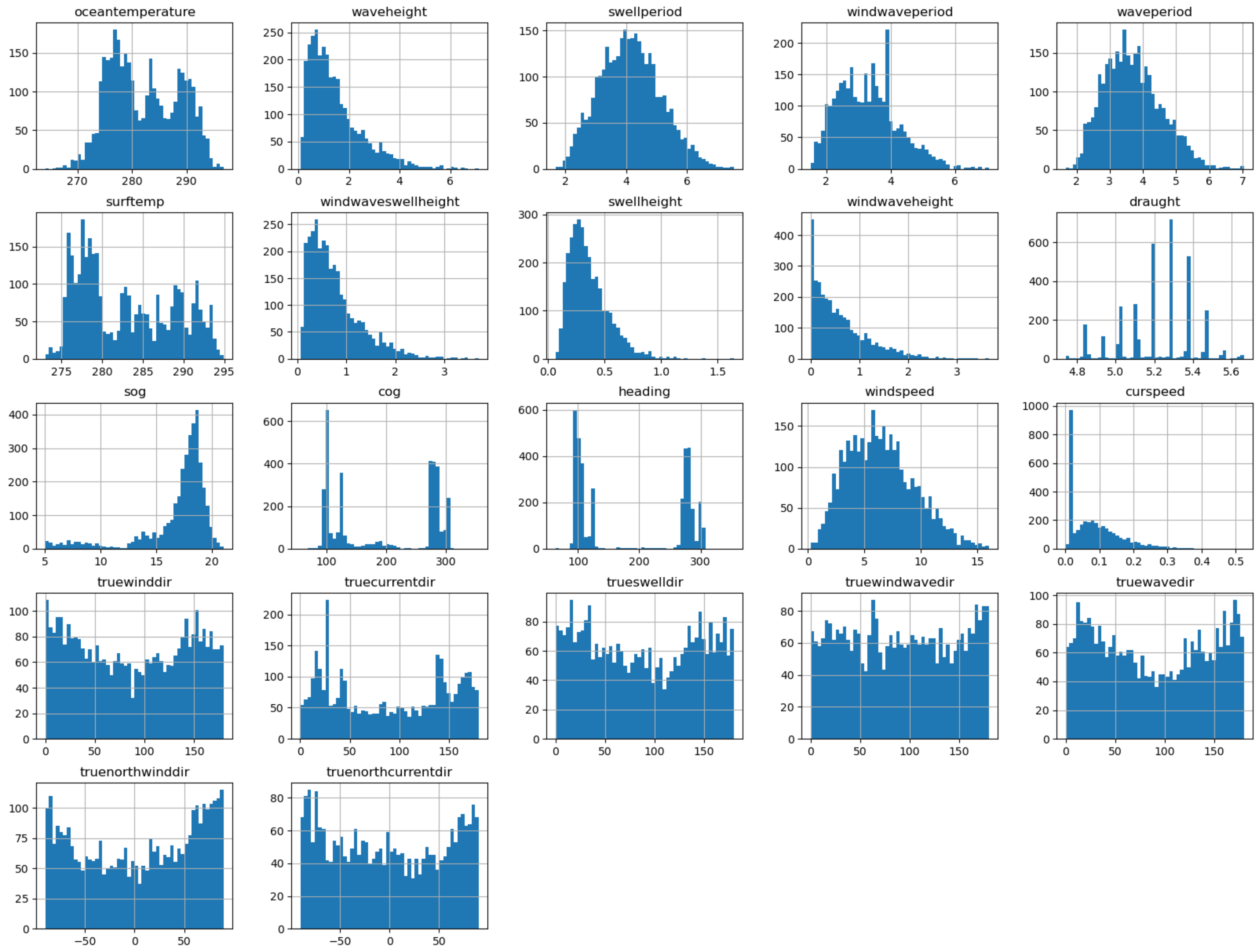

Figure 1: Histogram of the features

- The correlation of the features against SOG are determined. It is found that :
  - Draught
  - Course Over Ground (COG)
  - heading
  - Wind Speed
  - Current Speed
  - True Current direction

  Have relatively stronger correlation to SOG compared to other features, albeit the correlation is a weak one
- The correlation between the features is displayed using the following the heat map. From the heat map it can be observed that between these features:
  - Waveheight and wind wave swell height
  - Waveheight and wind wave height
  - Windwaveswellheight and wave period

  Have a strong correlation between each other.
- Open topic:
  - Feature reduction is possible, [1] suggested high feature correlation filter, the filter suggest that two features which has a high correlation (> 90%) is to be combined into a single feature. But the author is unsure whether this combination is physically sensible. Hence, this filter is yet to be applied for feature reduction.
  - Some of these features can be connected through wave equations, but the author has not found an equation which could relate these features.
- The random forest regressor could not function when `NaN` values are present. With that, the missing values are filled in using the `imputer` function. The missing values are filled in by means of `KNN`.

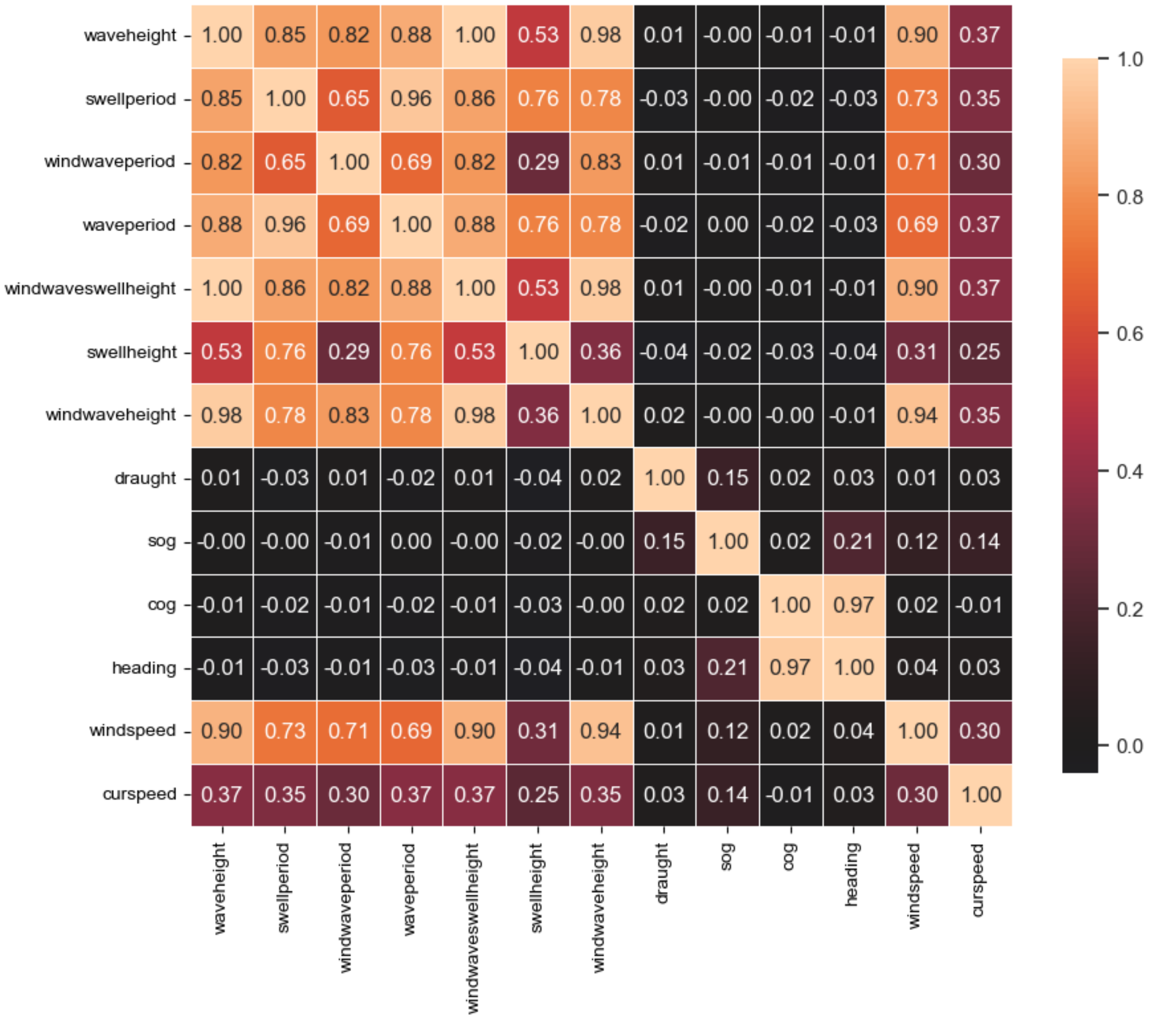

Figure 2: Correlation Heat Map

3.3 Modelling

- The data is split into 80:20 ratio. But considering the validation data, it is split into approximately 73:18:9.

- The model is then trained using Random Forest Regression (RFR). Additional training is also performed using Decision Tree Regressor (DTR). DTR model performance will be used as a benchmark as it is also a tree-based modelling method with similar methodology to RFR.

- The computational time of DTR is significantly faster than RFR Model Evaluation

3.4 Predicting Ship's Speed Through Water (STW)

- The ship's STW can be calculated using vector component of the SOG and current speed. The direction used will be according to True North. [4, 5]

- SOG represents the speed of the ship with reference to the ground, while the STW represent the ship's speed with reference to water.
- SOG also can be termed by the ship's speed that is captured by the GPS, and does not consider any effect of the current
- This means that the ship's STW will be greater than the ship's SOG when there is current moving against the ship's movement direction and vice versa
- The vector decomposition can be defined from the following equations, which is based on the equation by [4]:
  - The ship's SOG V_g can be decomposed into V_g^x and V_g^y, which represents the x and y components of the SOG respectively using the ship's course heading (COG) β *with respect to True North*:

$$V_g^x = V_g \sin(\beta) \tag{1}$$

$$V_g^y = V_g \cos(\beta) \tag{2}$$

  - To consider the effect of sea current. The current speed V_c will also be decomposed to x and y components respectively using the current direction γ *with respect to True North*:

$$V_c^x = V_g \sin(\gamma) \tag{3}$$

$$V_c^y = V_g \cos(\gamma) \tag{4}$$

  - from here the ship' STW V_{wx} and V_{wy} component can be found from the following equation:

$$V_w^x = V_g^x - V_c^x \tag{5}$$

$$V_w^y = V_g^y - V_c^y \tag{6}$$

  - The magnitude of the STW can be readily obtained from the following vector synthesis

$$V_w = \sqrt{(V_w^x)^2 + (V_w^y)^2} \tag{7}$$

- This principle is applied to the following Python script.

```
import np
dfprog["vgms"] = dfprog["sog_pred"]/1.9438
# HELP !
dfprog["vgx"] = dfprog["vgms"] * np.sin(np.deg2rad(dfprog["cog"]))
dfprog["vcx"] = dfprog["curspeed"] * np.sin(np.deg2rad(dfprog["
    truenorthcurrentdir"]))
dfprog["stw_x"] = (dfprog["vgx"] - dfprog["vcx"])

dfprog["vgy"] = dfprog["vgms"] * np.cos(np.deg2rad(dfprog["cog"]))
dfprog["vcy"] = dfprog["curspeed"] * np.cos(np.deg2rad(dfprog["
    truenorthcurrentdir"]))
dfprog["stw_y"] = (dfprog["vgy"] - dfprog["vcy"])

dfprog["vwms_p"] = np.sqrt(dfprog["stw_x"]**2 + dfprog["stw_y"]**2)
dfprog["stw_pred"] = dfprog["vwms_p"]*1.9438
nd{lstlisting}
```

4 Result and Discussion

The result of the research is discussed in this chapter. This comprises model validation and how different statistical metrics are used to analyze the model's performance.

4.1 Model Evaluation

The model are tested against four metrics, namely:

- R^2 : Indicate model fit. Best Score = 1
- Explained Variance `EV` : Indicate amount of variance in model. Best Score = 1
- Mean Absolute Error `MAE` : Indicate how much error a model makes in its prediction. Best Score = 0
- Root Mean Square Error `RMSE` : Same as MAE, more sensitive to outlier. Best Score = 0
- Median Absolute Error `MAD` : Check robustness against outlier. Best Score = 1

The result is summarized in the following table

Model	RFR	DTR
R^2	0.9328181446941499	0.8526085810220092
EV	0.932872958708872	0.8526260247615258
MAE	0.5546347329650284	0.8108982427834758
RMSE	0.7095480848510665	1.5566896535262504
MAD	0.3848463591000087	0.5475717149999983

Table 1: Model performance

Listing 1: Model Evaluation

```
def predict_y(x_test,model_type):
    y_predicted = model_type.predict(x_test)
    return y_predicted

def display_scores(x_test,y_test,model_type):
    from sklearn.metrics import explained_variance_score,mean_absolu
    y_predicted = model_type.predict(x_test)
    print(”R^2␣score␣(Indicate␣model␣fit.␣Best␣Score␣=␣1):”, model_t
    print(”Explained␣Variance␣EV␣(Indicate␣amount␣of␣variance␣in␣mo
    print(”Mean Absolute Error MAE (Indicate how much error a model
    print(”Root Mean Square Error RMSE (Same as MAE, more sensitive
    print(”Median Absolute Error MAD (Check robustness against outli

y_predicted = predict_y(x_test,model_rfr)
display_scores(x_test,y_test,model_rfr)
```

5 Summary and Outlook

In this chapter the summary of this research will be discussed. This section includes reflections of the research process and presents any possible suggestions and recommendations in this line of research. This chapter concludes this thesis.

References

[1] Misganaw Abebe, Yongwoo Shin, Yoojeong Noh, Sangbong Lee, and Inwon Lee. Machine learning approaches for ship speed prediction towards energy efficient shipping. *Applied Sciences*, 10(7):2325, 2020. 3, 5, 7

[2] Christos Gkerekos, Iraklis Lazakis, and Gerasimos Theotokatos. Machine learning models for predicting ship main engine fuel oil consumption: A comparative study. *Ocean Engineering*, 188:106282, 2019. 3, 5

[3] E. Bal Beşikçi, O. Arslan, O. Turan, and A. I. Ölçer. An artificial neural network based decision support system for energy efficient ship operations. *Computers & Operations Research*, 66:393–401, 2016. 5

[4] Liqian Yang, Gang Chen, Jinlou Zhao, and Niels Gorm Malý Rytter. Ship speed optimization considering ocean currents to enhance environmental sustainability in maritime shipping. *Sustainability*, 12(9):3649, 2020. 5, 8, 9

[5] Yang Zhou, Winnie Daamen, Tiedo Vellinga, and Serge P. Hoogendoorn. Impacts of wind and current on ship behavior in ports and waterways: A quantitative analysis based on ais data. *Ocean Engineering*, 213:107774, 2020. 8