Webinar on

Development of a Fuel Efficiency Prediction Model using Ship in-Service Data – A Machine Learning Approach

O P Sha

Department of Ocean Engineering and Naval Architecture, IIT Kharagpur

Organised by

FROST, Bhubaneswar

Saturday, 31 July 2021

CONTENT

- Introduction – EEDI SEEMP EEXI
- Data Science, AI and Machine Learning
- Machine Learning Tasks, Approach and Algorithms
- Fuel efficiency Prediction – Machine Learning Model
- Results
- Observation & Conclusion
- Some Examples of Machine Learning Application

Introduction

Environmental pollution, including GHG emissions, is a major concern in shipping.
IMO initiated the following conventions to overcome environmental challenges

1. Marine Environmental Protection Committee (MEPC)
2. International Convention for the Prevention of Pollution from Ships (MARPOL)

MARPOL Annex VI adopted (IMO, 2011b) to prevent GHG emissions by limiting ship exhaust gases CO_2, SO_x, NO_x and other ozone depleting substances.
Main contributor in GHG emissions is CO_2.

Appropriate CO_2, SO_x and NO_x reduction measures to be introduced to optimize the fuel consumption under various energy efficient navigation conditions (i.e. weather routing, optimal trim, appropriate speed) in ships.

Furthermore, various advancements in energy efficient technology to reduce the fuel consumption in marine engines have been considered in recent years

Introduction – Energy Efficiency Measures

Energy Efficiency Design Index (**EEDI**) – new designs to meet reference level of their ship type

- Aims at promoting efficient (less polluting) equipment and engines
- Requires minimum energy efficiency level per tonne-mile

Ship Energy Efficiency Management Plan (**SEEMP**) – an operational measure

- Establish a mechanism to improve the energy efficiency of the ship in cost-effective manner
- Uses a monitoring tool called Energy Efficiency Operational Indicator (EEOI)

Energy Efficiency Existing Ship Index (**EEXI**) –to reduce GHG emission by 40% (2030) 50% (2050)

- To apply to all vessels > 400GT from 2023
- Required EEXI ≈ EEDI for new ships as of 2023
- Ships rated on a scale of A to E based on annual operational carbon intensity indicator (CII)

Introduction – Objectives of the present study

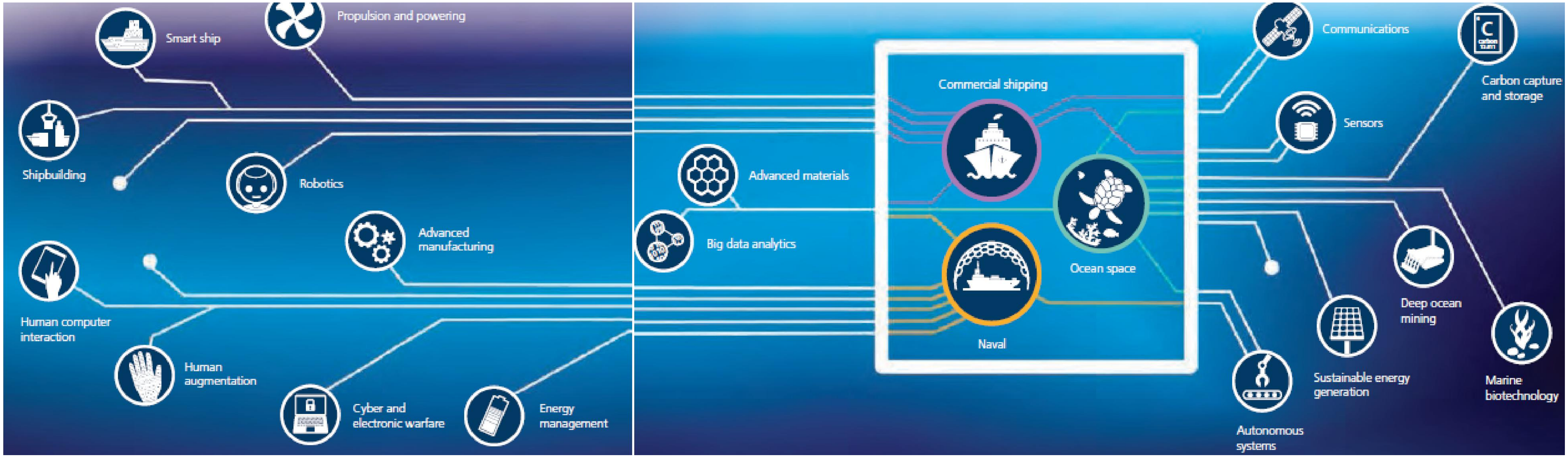
Statistical data analyses on several navigation parameters of a selected vessel are considered

- wind speed and direction
- average draft and trim,
- Speed over ground SOG
- log speed,
- ME power,
- shaft speed and
- fuel consumption

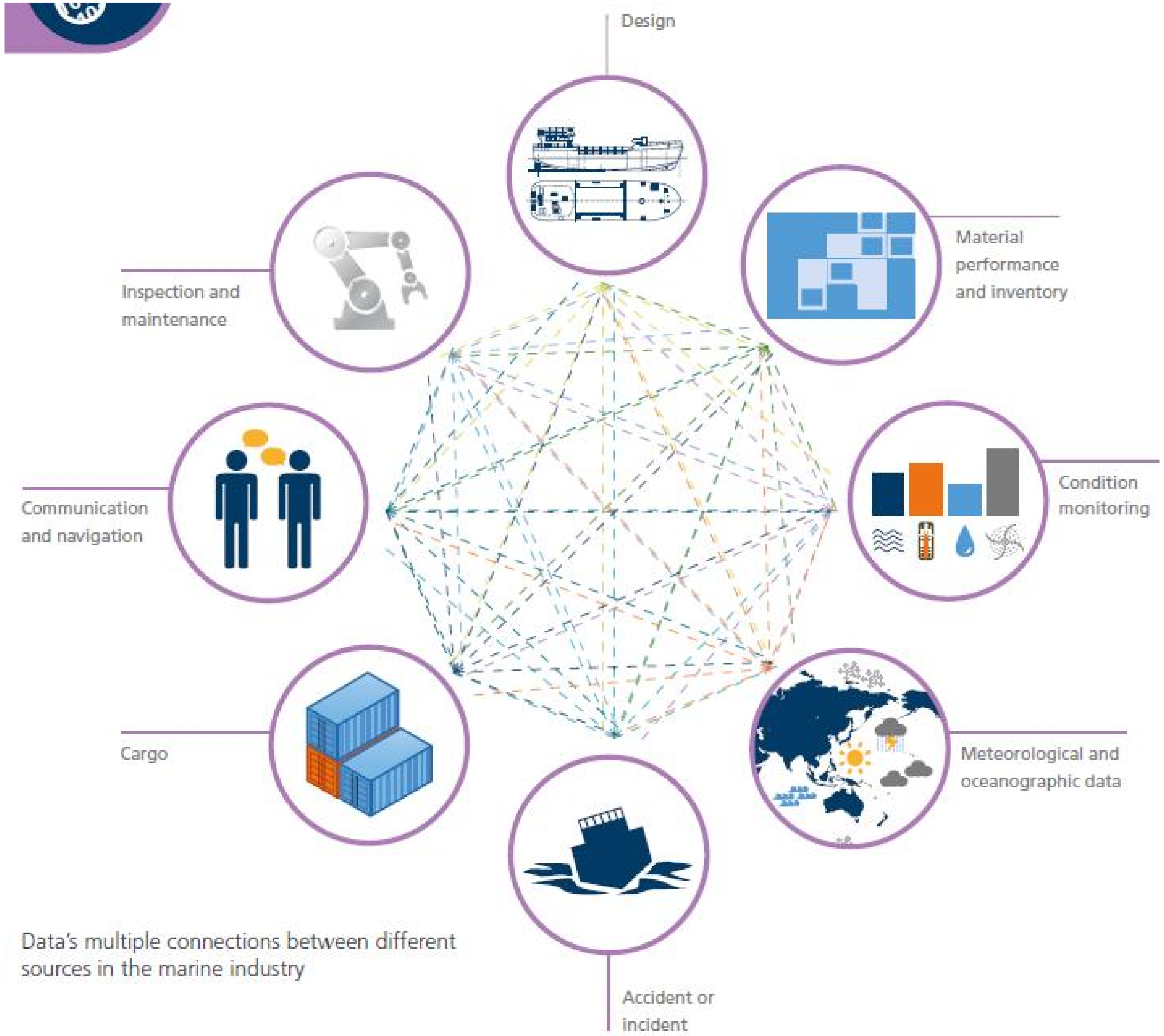
✓ Higher fuel consumption trends under these parameter variations to be noted.
various ship trim configurations with respect to the main engine (ME) fuel consumption rates are analysed to identify the optimal navigation conditions for the respective draft values of the ship.

This analysis can be an important part of the Ship Energy Efficiency Management Plan (SEEMP) and EEXI and that is the main contribution of this study.

Interrelationship between Technologies and Maritime Sectors
[Global Marine Technology Trends 2030]



Big Data Analytics [Global Marine Technology Trends 2030]



Data's multiple connections between different sources in the marine industry

Data Science – Artificial Intelligence AI

Data science (Wikipedia) is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains.

- Data science comprises of various statistical techniques.
- With data science we build models that use statistical insights

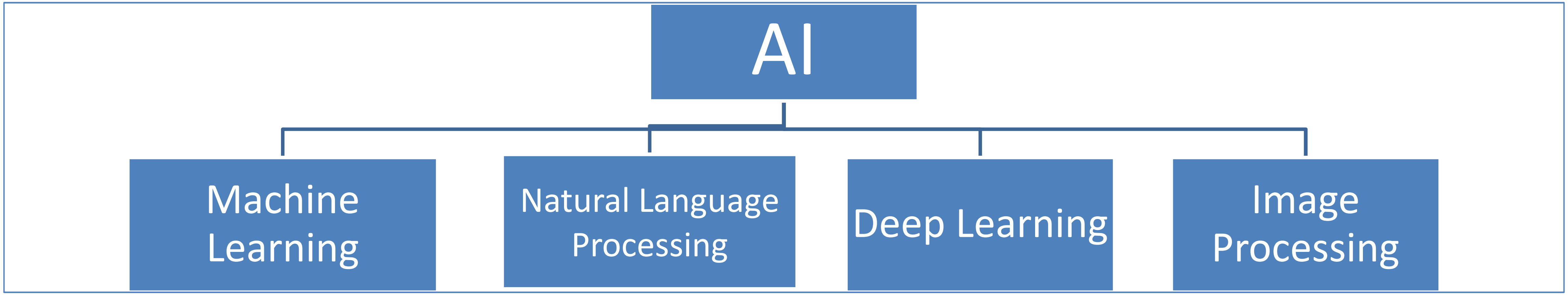
AI will make computer algorithms to solve problems by imparting autonomy to data model

- AI is for building models that emulate cognition and human understanding.

AI word was coined by Prof John McCarthy at the Dartmouth conference in 1956

Defn: *Algorithms enabled by Constraints exposed by Representation that support Models targeted at Thinking, Perception, Action (Prof Patrick Winston MIT AI Lab)*

Artificial Intelligence(AI) – Machine Learning (ML)



Machine Learning :

Field of study that gives computers the ability to learn without being explicitly programmed –Arthur Samuel (1959)

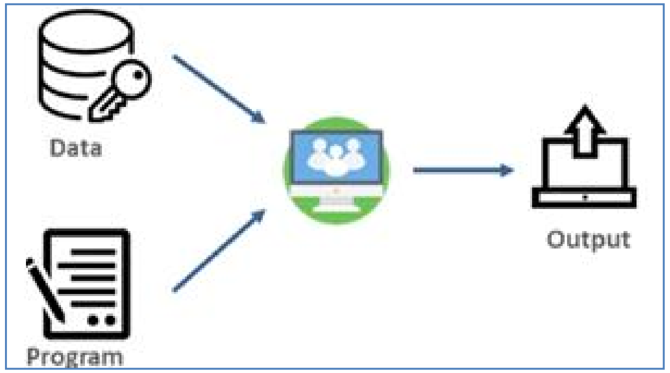
Machine learning is a method of data analysis that automates analytical model building

Need for Machine Learning:

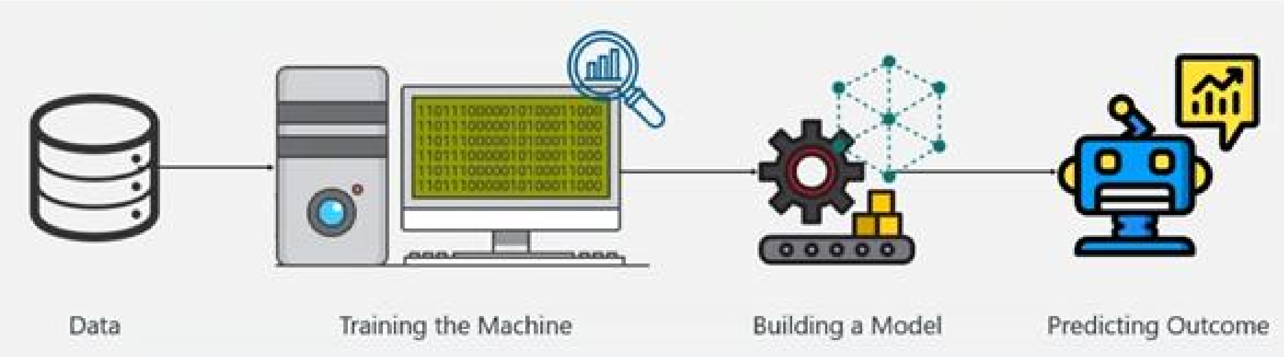
- Increase in data generation
- Uncover patterns and trends in data
- Solve complex problems
- Improve decision making

Machine Learning (ML)

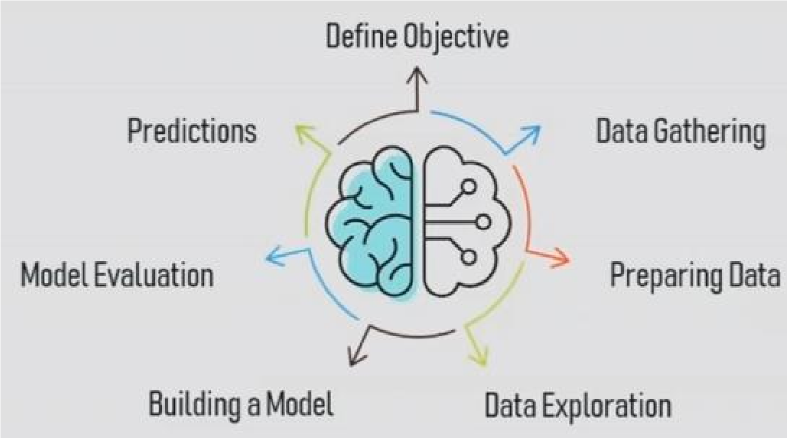
Traditional Programming



Machine Learning



Machine Learning Processes - involves building a predictive model that can be used to find solution for a problem statement



1. Define objective – target & predictor variables, - regression, binary, clustering
2. Data gathering
3. Data preparation – discover, format, clean, enrich
4. Exploratory Data Analysis(EDA) – understand trends and pattern in data
5. Building a Machine Learning Model – predictive model is built using ML algo
6. Model Evaluation and Optimisation -
7. Prediction

Machine Learning

Machine Learning Tasks

Supervised learning

Technique in which we teach or train the machine using data which is well labelled

Unsupervised learning

Training of a machine using information that is unlabelled and allowing the algorithm to act on that information without guidance.

Discovering hidden patterns in data

Reinforcement learning

Part of machine learning where an agent is put in an environment and he learns to behave in this environment by performing certain actions and observing the rewards which it gets from those actions.

Machine Learning Approaches

Types of machine learning problems can be classified into 3 types:

- Regression
  - ✓ Supervised learning
  - ✓ Output is continuous quantity
  - ✓ Main aim is to forecast or predict
- Classification
  - ✓ Supervised learning
  - ✓ Output is a categorical quantity
  - ✓ Main aim is to forecast or predict
- Clustering
  - ✓ Unsupervised learning
  - ✓ Assign data points into clusters
  - ✓ Main aim is to group similar items into clusters

Machine Learning

Supervised Learning Algorithms:

- Linear Regression
- Logistic Regression
- Decision Tree
- Random Forest
- Naïve Bayes Classifier
- K- Nearest Neighbour (KNN)
- Support Vector Machine (SVM) Non-linear SVM

Supervised Learning Algorithms

Linear regression: Method to predict dependent variable (Y) based on values of independent variables (X)

Used for cases where we want to predict some continuous quantity

Predictor variable = X;   Response variable = Y

Logistic regression: Method to predict a dependent variable based, given a set independent variables

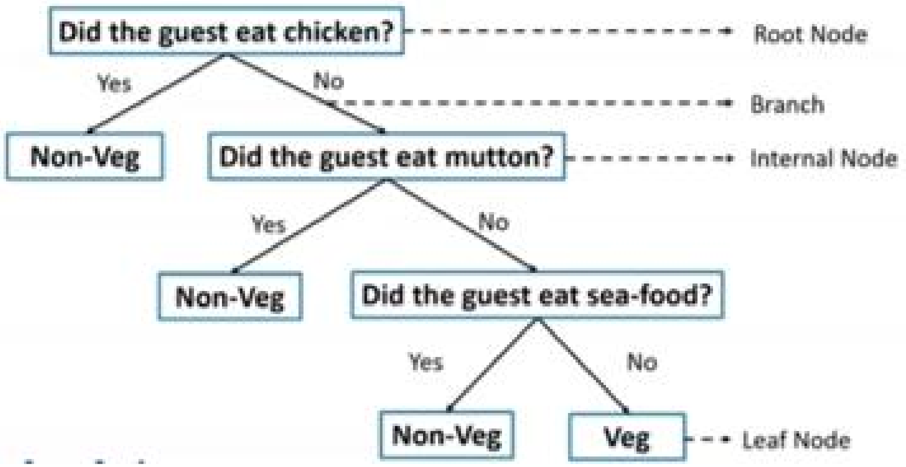
The dependent variable should be categorical

Decision tree: Supervised Machine Learning Algorithm which looks like an inverted tree

Node represents predicator variable (feature)

Link between nodes represents decision

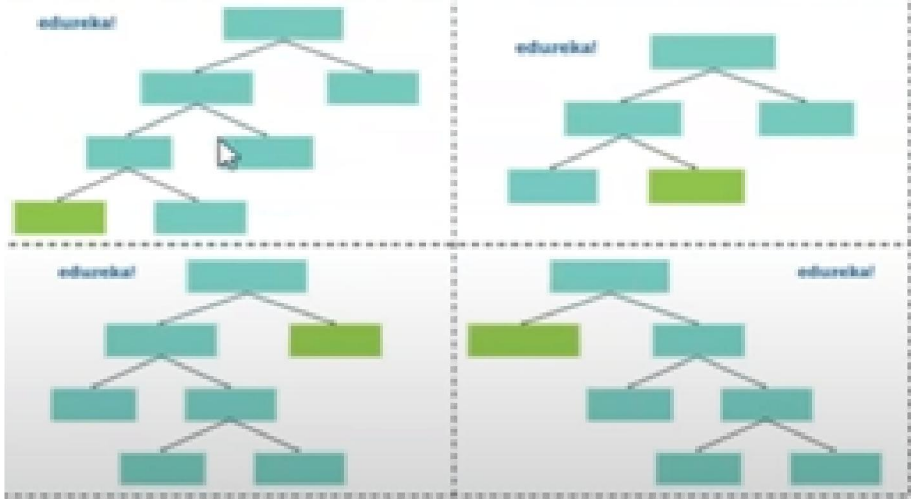
Leaf node represents an outcome (response variable



Supervised Learning Algorithms

Random Forest: Builds multiple decision trees - glues them together to get a more stable & accurate prediction

Each decision tree predicts the output class based on the respective predicator variables used in that tree



Naive Bayes: Supervised classification algorithm based on Bayes theorem

Used for solving classification problems following a probabilistic approach

Here the predicator variables in the ML model are independent of each other

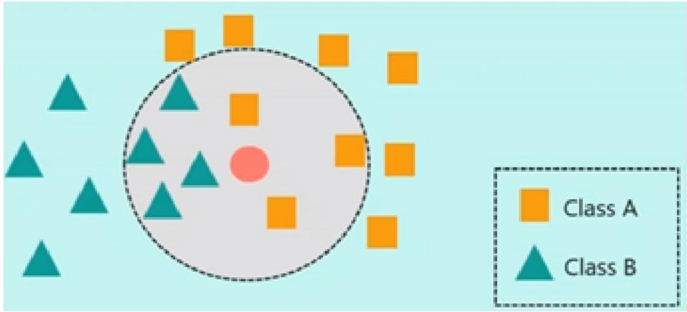
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Supervised Learning Algorithms

K Nearest Neighbour: Supervised classification algorithm (KNN)

Classifies new data point into target class, depending on the features of its neigbouring data points

For K = 7, find the nearest neighbours

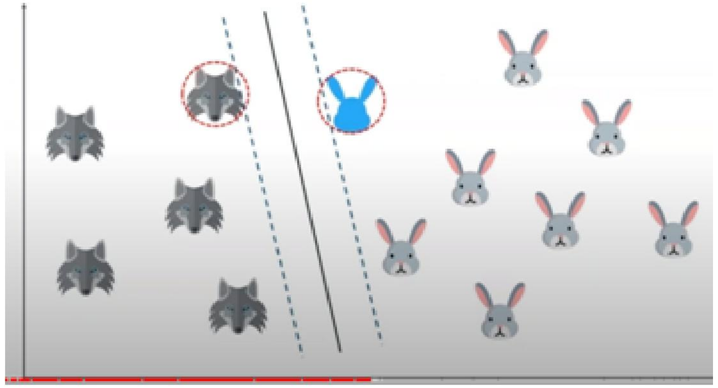


Ans = Class B

Support Vector Machine: Supervised classification algorithm (SVM) – Classification and regression algorithm

Method that separates data using hyperplanes

SVM Kernel functions



Best hyperplane that has the maximum margin

Architecture of a Multilayer Perceptron - ANN

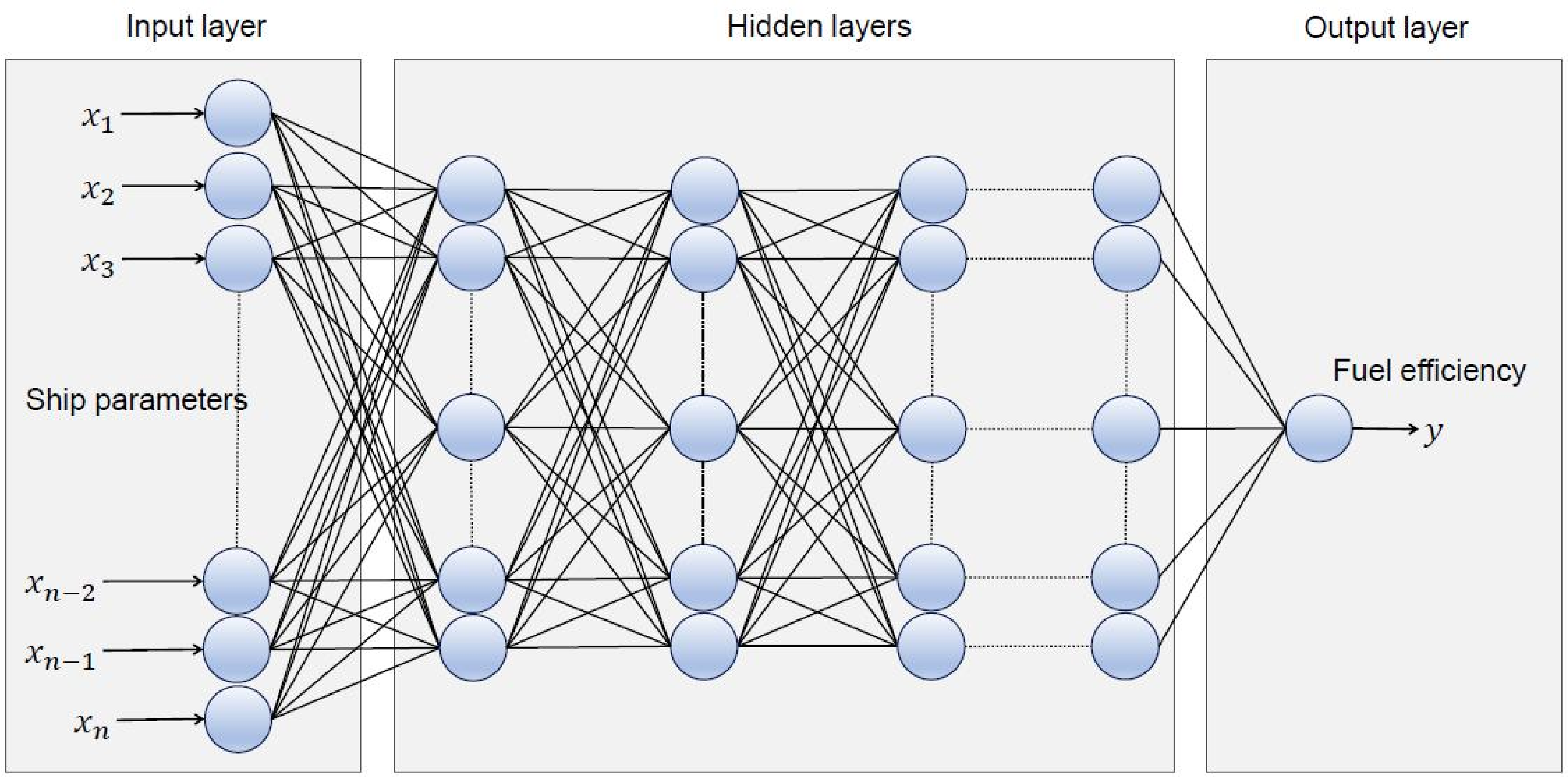
ANN is a simplified model of the human neural network structure, which sums the products of the input chemical signal coming from the synapse, with the weight of the connection strength of the synapse, and finally the output value is derived via an activation function. A simple ANN is expressed by

$$v = b + \sum_{i=1}^{n} w_i x_i$$

$$y = f(v)$$

$$f(x) = max(0, x)$$

x_i = i^{th} input
w_i = i^{th} weight
b = bias
v = summed output
f = activation function
y = output



CASE STUDY – MARORKA DATASET

<u>Data Gathering</u>

Dataset used for analysis is collected from a sensor-based monitoring system called Marorka which picks up signals from various measuring devices like

•shaft power meter, • fuel flow meter, • anemometers, etc.

every 5 seconds which in turn is accumulated/averaged every 15 minutes.

This is then transmitted on-shore through their servers.

Dataset provided by the shipping company for their bulk carrier contains data of 15 minute time stamps from 21 Nov 2014 to 31 Dec 2015

DATASET

Data Preparation

The parameter "relative wind direction" is given in terms of angle ranging from 0 to 359 degrees – this results in circular discontinuity. Therefore the wind direction (θ in degrees) along with the relative wind speed (V_{wind} in m/s) is used to create two new features:

$$V_{headwind} = V_{wind} \times \cos(\theta_{wind})$$

$$V_{crosswind} = V_{wind} \times \sin(\theta_{wind})$$

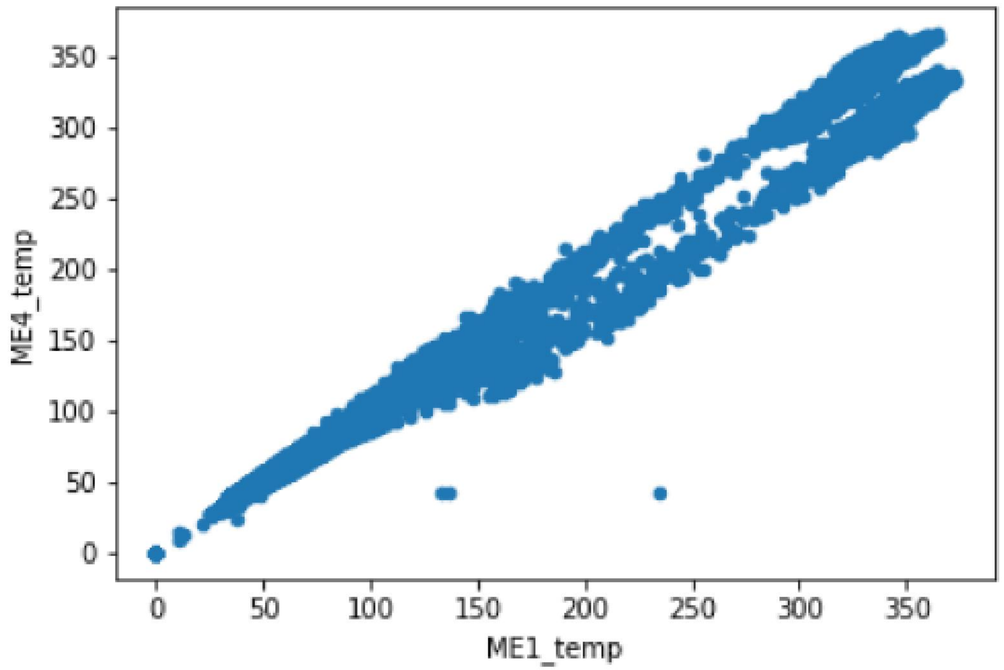
The parameters draught-fwd (T_{fwd}) and draught-aft (T_{aft}) are used two create two new features;

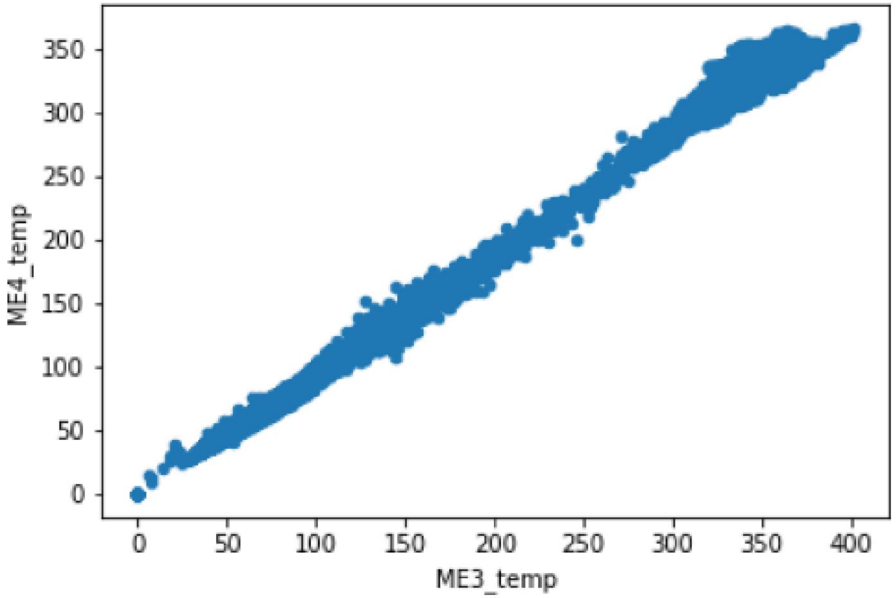
$$\text{trim} = T_{fwd} - T_{aft}$$

$$\text{draught} = 0.5 \text{ X } (T_{fwd} - T_{aft})$$

Plots showing correlation between Main Engine Exhaust Gas Temperatures

Main Engine 6 cylinder

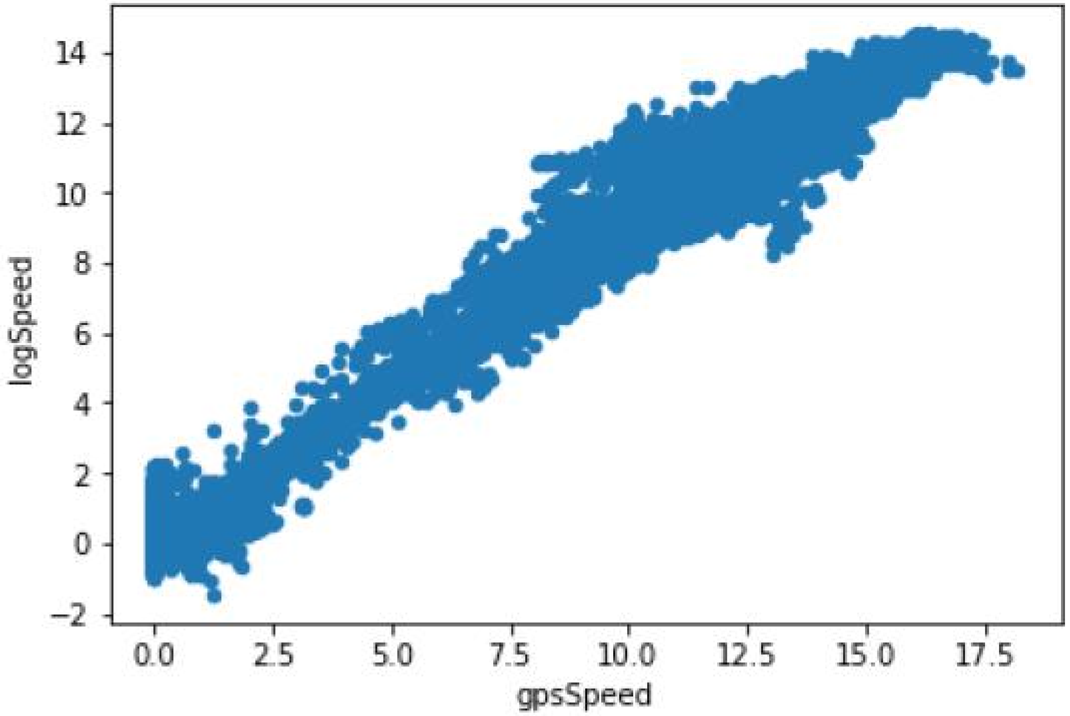




It is assumed that this temperature can be an indicative of the engine’s performance.

Since the values of these temperatures are highly correlated, as seen in the above plots, the average of the six cylinders is being used as a new feature.

Plots showing correlation between logSpeed and gpsSpeed



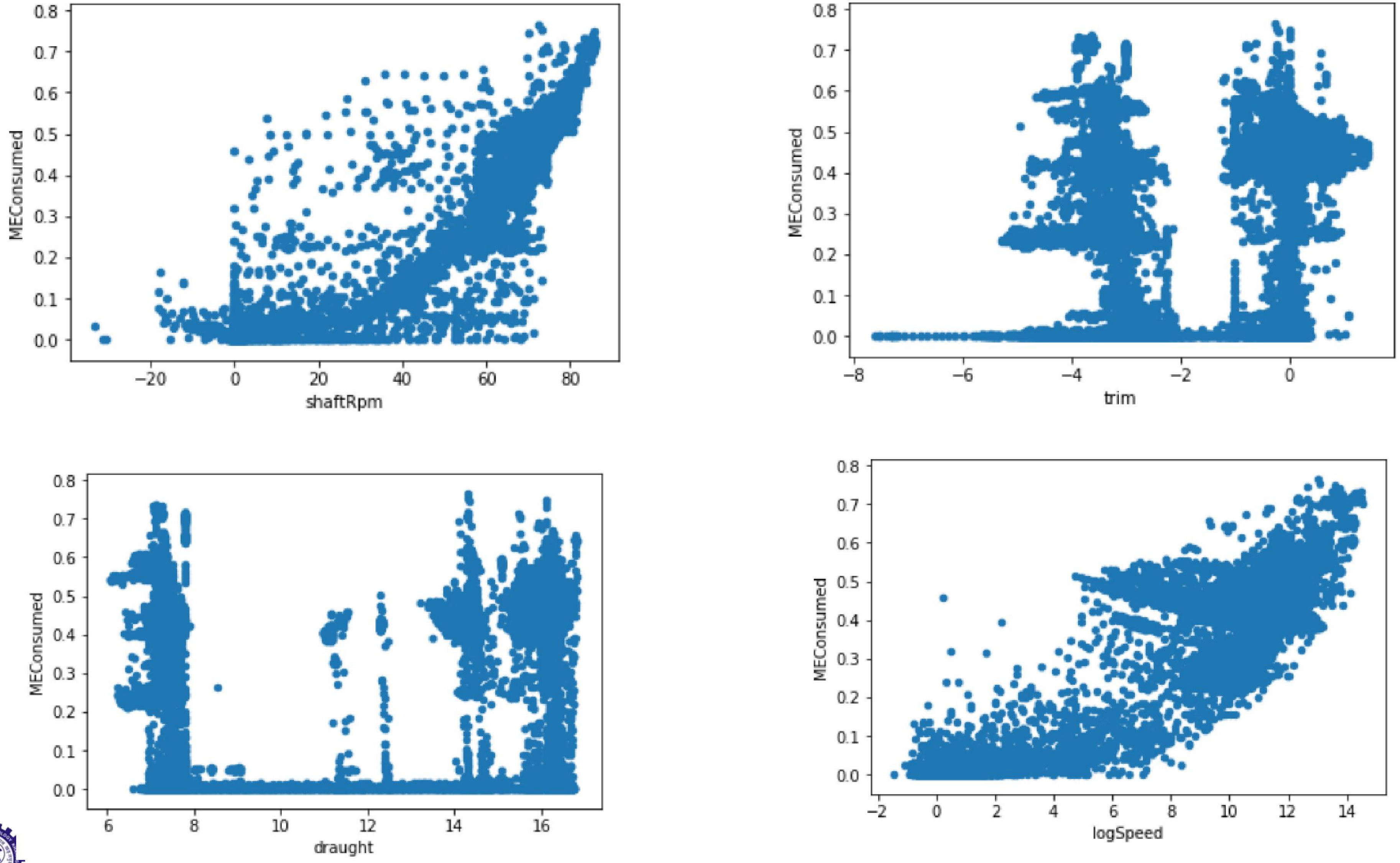
From the above plot, logSpeed and gpsSpeed are highly correlated
hence only one of them. i.e. logSpeed has been used for carrying out the analysis

Description of Predicator and Response variables

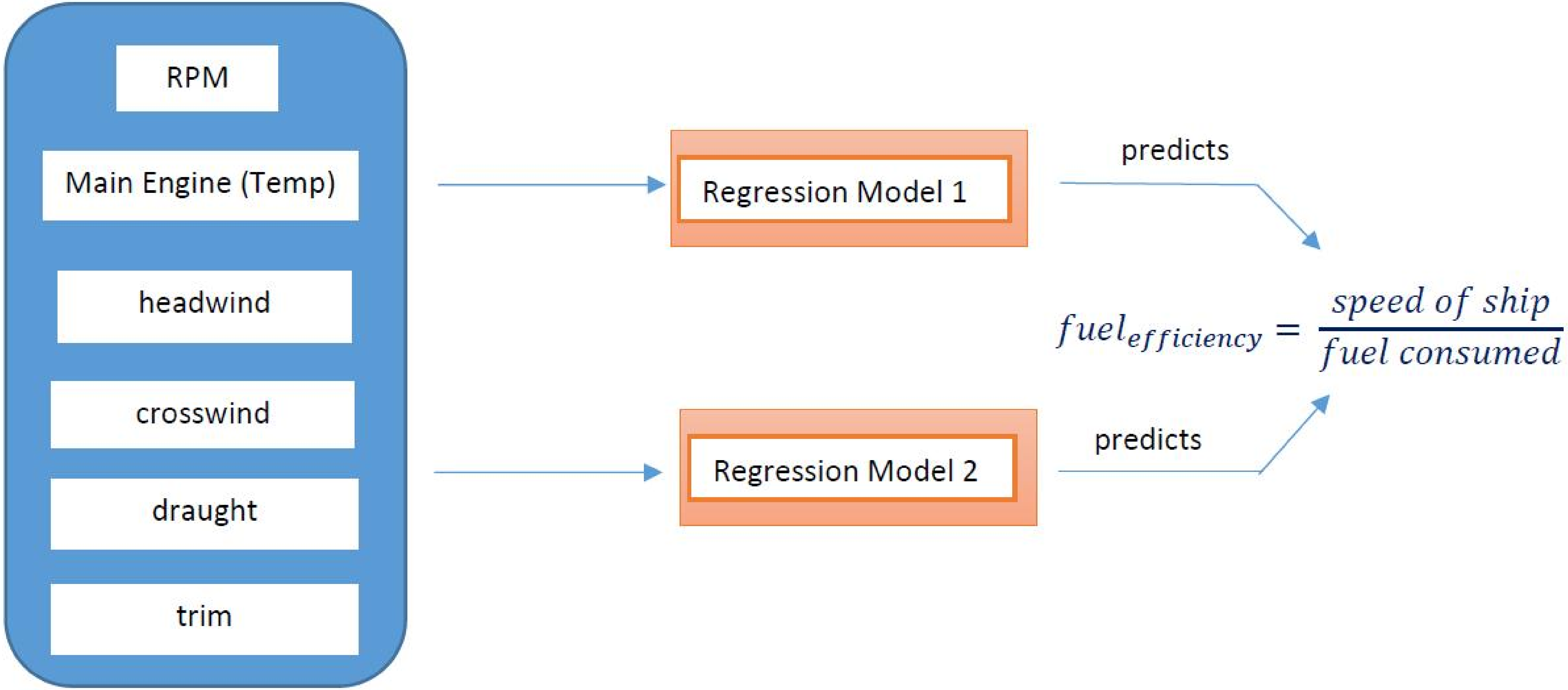
Sl No	Parameters	Nomenclature	unit	Data type
1	Fuel consumed by main engine	MEConsumed	tonnes	output
2	Propeller shaft rpm	shaftRpm	rpm	input
3	Speed of ship from log	logSpeed	knots	output
4	Main Engine average temperature	ME_temp	degree Celsius	input
5	Head wind	headwind	metres/sec	input
6	Beam wind	crosswind	metres/sec	input
7	Mean draught	draught	metres	input
8	Trim by bow positive	trim	metres	input

35344 data were retrieved from the Marorka data

Scatter plots showing correlation between fuel consumed and other variables



Methodology to Predict Fuel Efficiency



$$fuel_{efficiency} = \frac{speed\ of\ ship}{fuel\ consumed}$$

Prediction of Fuel Efficiency

Table 1: Input and output of regression model – fuel consumption

Input to the Model 1	Output
shaftRpm	MEConsumed
ME_temp	
headwind	
crosswind	
draught	
trim	

Table 2: Input and output of regression model – ship speed

Input to the Model 2	Output
shaftRpm	logSpeed
ME_temp	
headwind	
crosswind	
draught	
trim	

$$fuel_{efficiency} = \frac{ship\ speed}{fuel\ consumed}$$

Prediction of Fuel Efficiency

Dataset after pre-processing is randomly divided into two sets: 70% of data set for training of the model
30% to test performance of trained model

For comparison of regression models R^2 score is used R^2 is the goodness of fit and is a statistical measure of how well the regression predictions approximate the real data

$$R^2 = 1 - \frac{\sum (y - y_{model})^2}{\sum (y - y_{mean})^2}$$

$R^2 = 1$; indicate perfect fit

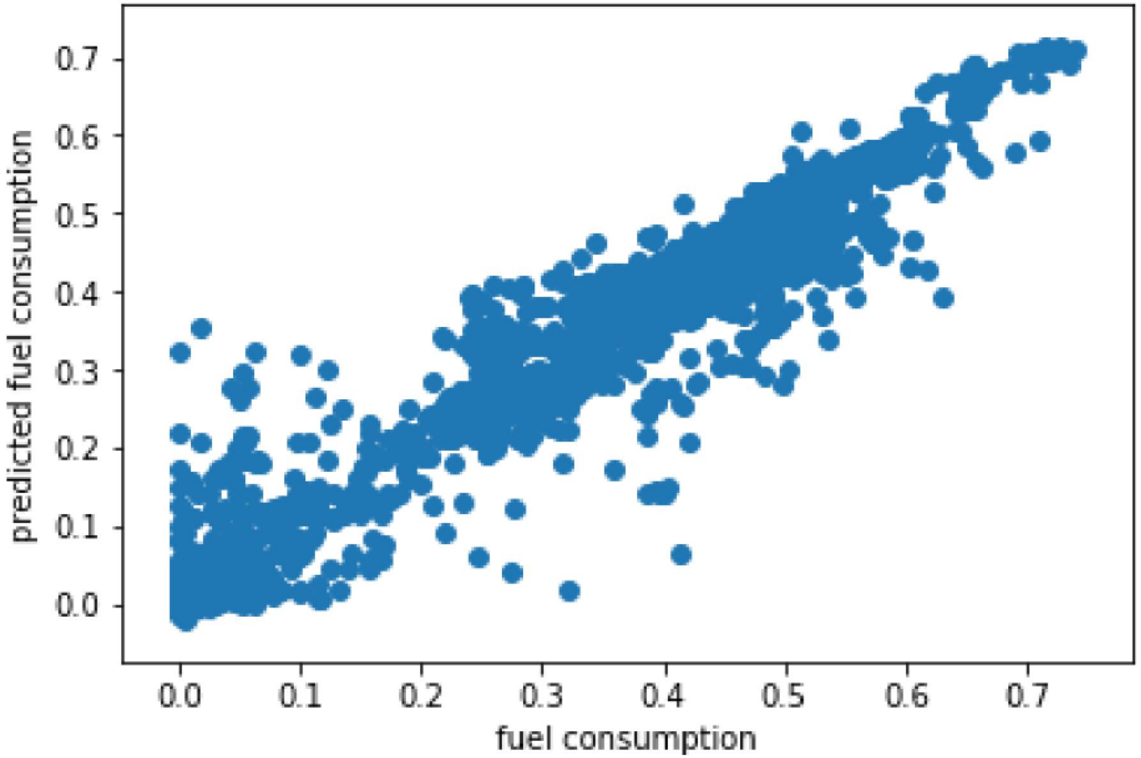
Results

Table 3: Results from various Machine Learning Models

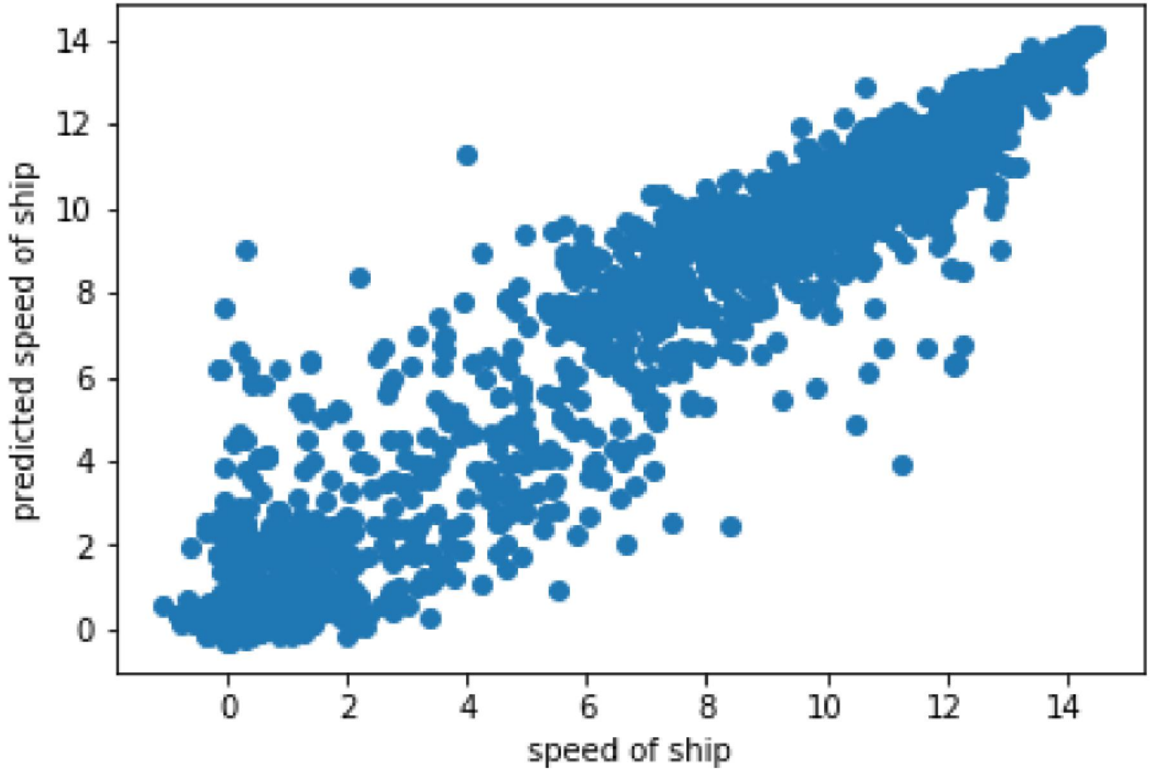
Regression Model	Fuel Consumption (MEConsumed)		Ship's speed	
	R^2	RMSE	R^2	RMSE
Linear Regression	0.857	0.063	0.922	1.043
Neural Network	0.881	0.057	0.944	0.882
Support Vector Machine	0.901	0.052	0.953	0.809
Gradient Boosting	0.970	0.029	0.965	0.699

From the above Table, it can be concluded that GRADIENT BOOSTING algorithm is able to predict the fuel consumption and ship speed with highest accuracy.

Results – Gradient Boosting Algorithm



Scatter plot: original and predicted fuel consumption



Scatter plot: original and predicted ship speed

Observations

A ship's state can be thought of consisting of:

1. **State variables:** – parameters that provides information about the ship's current state

   - shaft rpm, engine temperature, draught, speed, wind

2. Control variables: parameter(s) that can be controlled – trim of vessel

The idea behind the regression model is to suggest the value of the control variable that can result in increased fuel efficiency.

Example is to find optimal trim for the state of the ship and sea conditions

Present models that are developed for the prediction of fuel efficiency is that they all are instantaneous maps of the state parameters to fuel efficiency

- temporal variation of such state parameters are not used in the modelling
- the dynamic nature of ship operation is not taken into account

Conclusions

1. Mathematical model using machine learning techniques can be developed that predict fuel efficiency of a ship under different operating and environmental condition

2. Sensors onboard ships provide the data to the mathematical models

3. Different ships will be trained differently to provide ship-specific solutions

4. These mathematical models have added applications, like
   - Anomaly detection: – look for systematic bias in case of a damage
   - Data-mining state variables: - used to gain understanding of ship's operation
   - Benchmark modifications like fuel saving devices: compare efficiency before and after

Conclusions

5. Importance of dynamic nature of ship operation – a change in control variable (trim) to improve fuel efficiency will affect the state variables.

However, such a system will be based on the assumption that the change of trim will not affect other state variables (Drawback).

6. Dynamic models that take into effect the temporal changes of these state parameters due to change in control variable needs to developed for more robust system

Other Maritime Example of Machine Learning (ML)

Predict Corrosion Allowance for Ship Structural Members

Particulars of structural members

Structural Member	Age of Ship in years (AGE)	Original Thickness of the Structural Member in mm (OT)
Hold Frame	13	23
Side Shell	12	25
Inner Bottom Plate	21	36
Bulkhead (Floodable Hold)	16	19

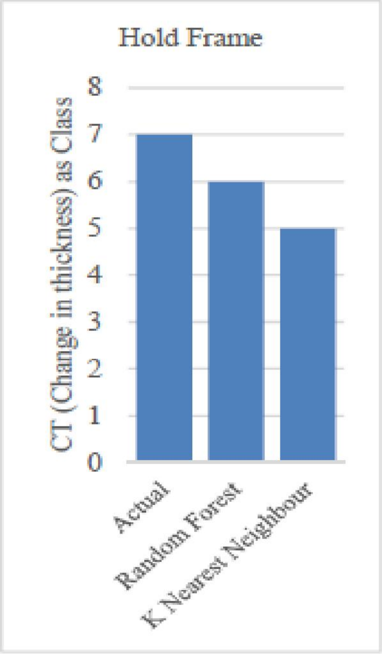


Figure 2. CT for Hold Frame

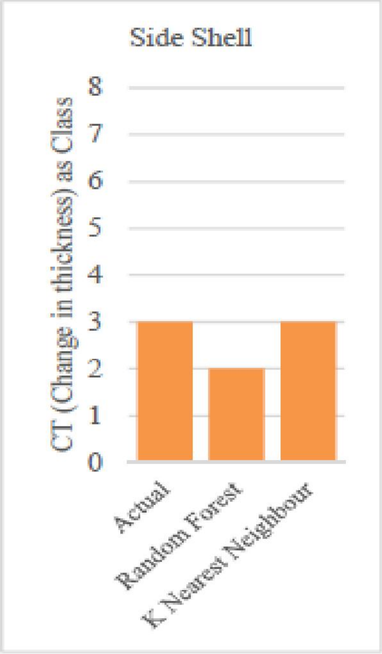


Figure 3. CT for Side Shell

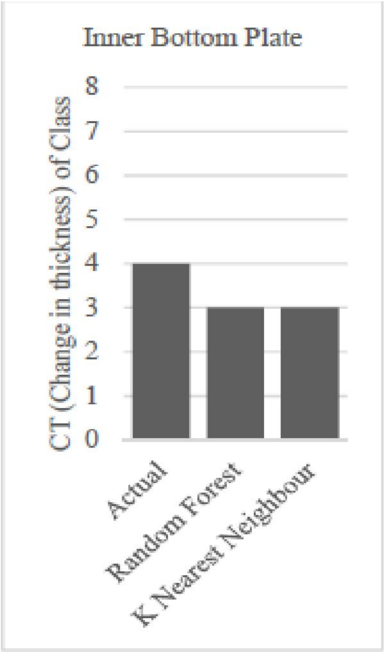


Figure 4. CT for Inner Bottom Plate

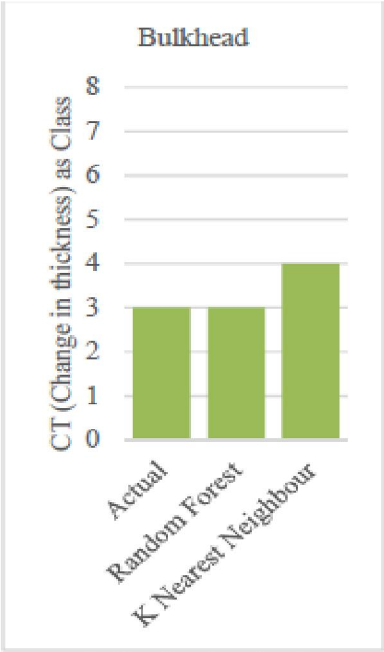
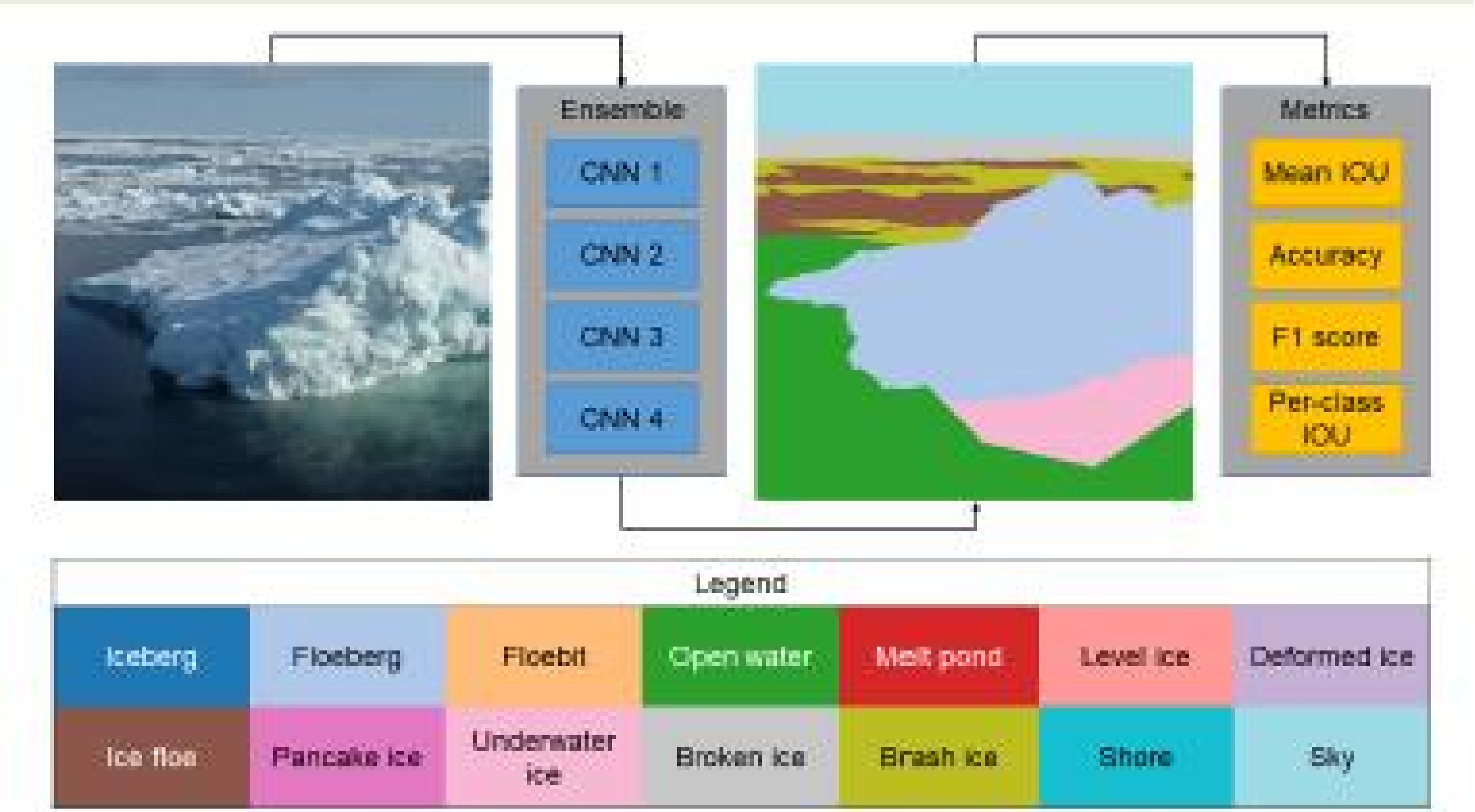


Figure 5. CT for Bulkhead

Other Maritime Example of Machine Learning (ML)

Supplementing remote sensing of ice: Deep Learning based image segmentation system for automatic detection of sea ice formation from close-range optical images



Thank You