



Seoul National University  
College of Engineering  
Department of Naval Architecture and Ocean Engineering  
1, Gwanak-ro, Gwanak-gu, Seoul 151-744, Korea

Fall 2020

머신러닝

PA # 6

Instructor name	김태완 교수님
Student name	이용준
Department	조선해양공학과
Student ID	2015-19595
Submission date	2020.12.23

# Contents

<b>1. Problem Definition</b>	<b>3</b>
<b>2. Problem Analysis</b>	<b>4</b>
<b>3. Code Explanation</b>	<b>5</b>
<b>4. Result &amp; Conclusion</b>	<b>5</b>
4.1 Result	5
4.2 Conclusion	8

## 1. Problem Definition

Input:

Column	Explanation
speed	선박 속도 (knot)
heading	진행 방향(radian)
rudder	러더 사용 값 (°)
Trim	트림(m)
forwtrd_Draft	선수 흘수(m)
after_draft	선미 흘수(m)
mid_port_draft	중양 좌현 흘수(m)
mid_starboard_draft	중양 우현 흘수(m)
wave_height	파고 높이(m)
wave_direction	파고 방향(radian)

Column	Explanation
wind_speed	바람 속도(m/s)
wind_direction	바람 방향(radian)
current_speed	조류 속도(m/s)
current_direction	조류 방향(radian)
pressure	압력(hPa)
water_temperature	수온(°C)
air_temperature	대기 온도(°C)
engine_rpm	엔진 rpm(r/min)
engine_pressure	엔진 압력(hPa)
engine_temperature	엔진 온도(°C)

Output:

Column	Explanation
power	추진 동력 (kN)

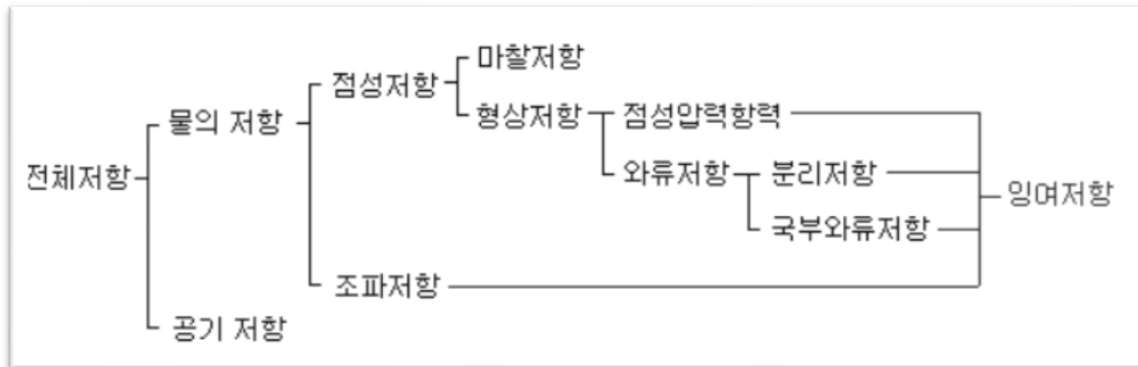
1. 모델 설계: 주어진 데이터로 추진 동력을 추론하는 회귀 모델을 만드는 것이 이번 과제의 목표이다. 이를 위해 주어진 Input data를 분석하여 필요한 데이터를 선택하여야 한다.
2. 모델 학습 및 분석: sklearn의 RandomForestRegressor를 활용하여 추정 모델을 학습시키고 결과를 확인한다. 이때 RMSE(root mean square error)를 통해 결과값을 비교하여 본다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

3. 정수 중 추진동력 추정: 학습한 모델을 사용하여 주어진 선박 속도에서 기상조건이 없을 때의 추진 동력을 추정해본다. 속도-추진동력 곡선을 그려보고 이를 주어진 곡선과 얼마나 유사한지 비교해본다.

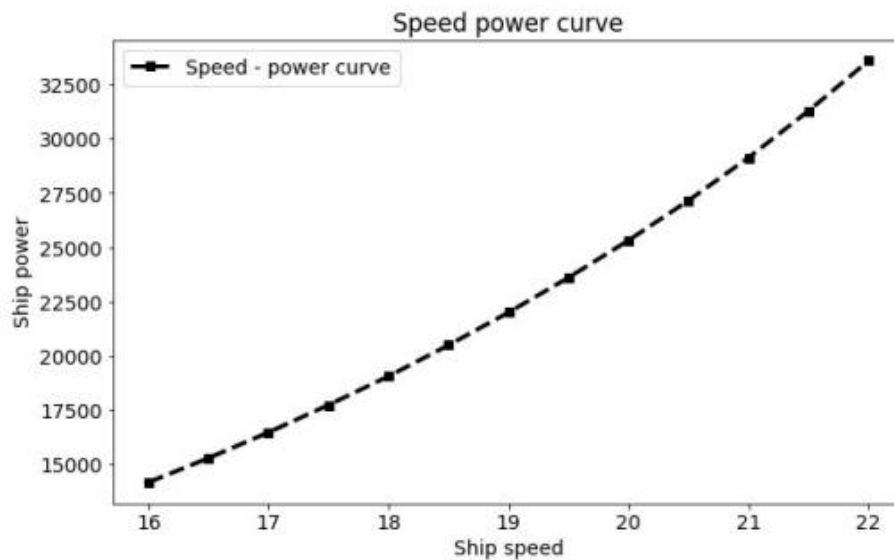
## 2. Problem Analysis

1. 입력 데이터를 결정하기 위해 선박이 받는 총 저항에 대하여 조사를 해볼 필요가 있다. 선박이 받는 저항은 다음과 같은 구성으로 이루어진다.



2. 선박의 속력과 동력은 다음과 같은 관계를 가지고 있다고 한다. 일반적으로 속력의 세제곱이 동력과 비례한다.

선박의 속력( $V$ )과 동력( $P$ )은 저속일 때  $V^{2.5} \sim P$ , 고속일 때  $V^{3.5} \sim P$  관계를 갖는다.



<속도-추진 동력 예시>

### 3. Code Explanation

프로젝트에 사용된 함수에 관한 설명이다.

RMSE(model, x, y_truth)		
Parameter	Type	Explanation
model	분류기	RandomForestRegressor 모델
x	df	학습 및 검증을 위한 Input data
y_truth	df	학습 및 검증을 위한 Output data

curve(x, coefficients)		
Parameter	Type	Explanation
x	array	정수 중 선박의 speed
coefficients	array	회귀 모델의 계수 정보들을 담고 있는 array

### 4. Result & Conclusion

#### 4.1 Result

**1번 문제** : 1번 문제에서는 조선공학의 관점에서 추진 동력과 연관되어 있는 입력 데이터를 선택하는 것이다. 각 항목에 대한 선택은 다음과 같다.

선박의 진행방향(heading)과 러더 값(rudder)은 선박 운동 조종에 대한 데이터로서 heading과 rudder는 선박의 운동 경로를 결정하는 대표적인 성분들이다. 선박이 직선 운동을 하는 지, 회전 운동을 하는 지는 추진 동력에 영향을 주기때문에 필요한 정보다.

트림과 흘수 정보들은 모두 배수량에 영향을 주는 성분들이다. 배수량은 선박의 조파저항과 매우 큰 연관을 갖고 있기때문에 추진 동력에도 영향을 미친다. 따라서 이 정보들은 필요한 정보로 판단하였다.

조류와 파고는 상황에 따라 선박의 동력을 크게 방해할 수 있는 데이터이기 때문에 중요한 데이터다. 바람 역시 선박에 공기 저항을 발생시키기 때문에 중요한 데이터다.

압력도 선체의 저항요소이고, 수온도 물의 점성에 영향을 주기 때문에 중요한 데이터라고 할 수 있다. 단, 대기 온도는 수온과 항상 비슷한 값을 가지고 있기 때문에 대기온도를 따로 고려할

필요는 없다고 생각을 하였다. 그래서 대기 온도 데이터는 제외하기로 하였다.

엔진 rpm은 추진 동력을 결정하는 데에 있어 중요한 요소다. 프로펠러의 회전 속도로 인해 추진력을 얻기 때문이다.

엔진의 압력(engine\_pressure)과 엔진의 온도(engine\_temperature)는 엔진의 성능, 즉 엔진 효율에는 영향을 주지만 엔진 rpm이 주어진 상황에서 이 데이터들은 선박의 추진 동력에는 무의미할 것으로 생각이 되어 이 데이터들은 제외하기로 결정했다.

결론적으로, 입력 데이터에서 제외하는 데이터는 대기 온도, 엔진 압력, 엔진 온도로 결정했다.

**2번 문제** : sklearn의 RandomForestRegressor 모델을 사용하여 데이터를 학습시키고 RMSE를 구해보았다. 결과는 다음과 같았다. 값이 그리 크지 않은 것으로 보아 웬만큼 학습이 잘 이루어졌다고 판단할 수 있다.

```
[RMSE]
train RMSE : 85.56177104869192
valid RMSE : 346.2676079945101
test  RMSE : 309.80506229760977
```

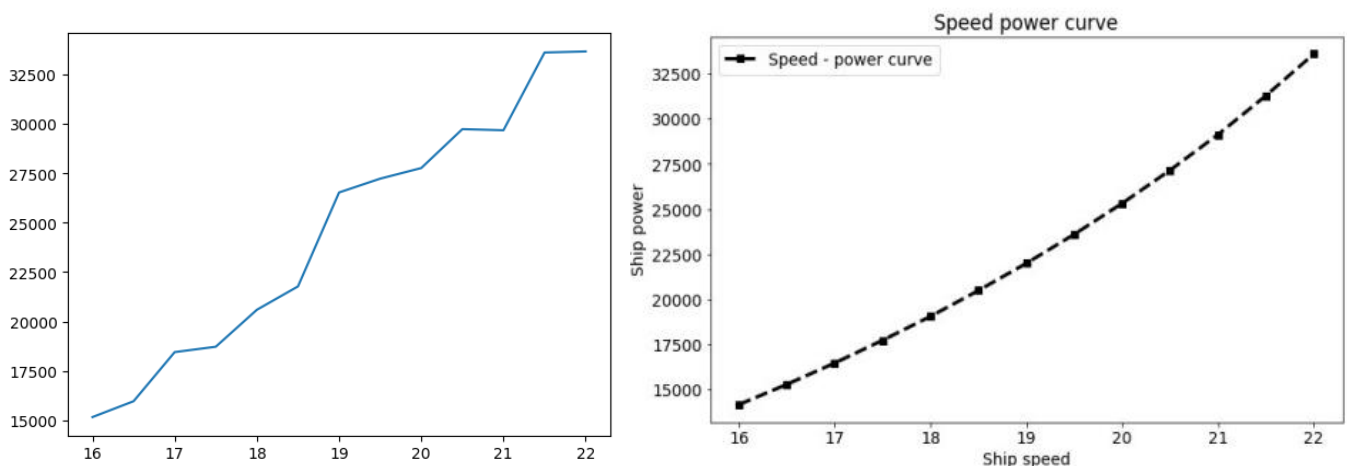
**3번 문제** : 학습 모델을 통해 정수 중 speed - power 곡선을 얻기 위해 입력 데이터를 csv파일로 새로 만들었다.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		speed	heading	rudder	trim	forward_dra	after_draft	mid_port	mid_starb	wave_heig	wave_direc	wind_spee	wind_direc	current_sp	current_dir	pressure	water_tem	engine_rpm	
2	0	16	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	55	
3	1	16.5	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	56.5	
4	2	17	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	58	
5	3	17.5	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	59.5	
6	4	18	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	61	
7	5	18.5	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	62.5	
8	6	19	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	64	
9	7	19.5	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	65.5	
10	8	20	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	67	
11	9	20.5	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	68.5	
12	10	21	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	70	
13	11	21.5	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	71.5	
14	12	22	0	0	0	12	12	12	12	0	0	0	0	0	0	0	0	73	
15																			
16																			

처음엔 입력 데이터 중 speed 성분만 16.0kn~22.0kn로 차등을 시키고 나머지 데이터는 모두 동일하게 입력을 하고 테스트를 해보았다. 그런데 결과가 매우 좋지 않았다. 이를 해결하기 위한 방법으로 모델을 학습할 때 각 입력 데이터 별 중요도를 산출해 비교해보았다. 결과는 아래와 같았다.

```
[feature importance]
speed      0.0005900552318193265
heading    0.0007709615316737091
rudder     5.7518477121538154e-05
trim       0.0015215073332487112
forward_draft  0.00154587184530544
after_draft  0.002247304620680914
mid_port_draft  6.93097581484247e-05
mid_starboard_draft  0.0006230721609959644
wave_height  0.002201136872557686
wave_direction  0.0003776491524553445
wind_speed  0.00228986161769418
wind_direction  0.0024547624091729573
current_speed  0.0002677544603405212
current_direction  0.000172354487325056
pressure    0.0104515187445374
water_temperature  0.0008359900306756189
engine_rpm  0.9735233712662472
```

그 결과 engine\_rpm의 feature importance가 0.97 정도라는 것을 발견했다. 이 모델은 engine\_rpm에 크게 의존한다는 것을 알게 되어 입력데이터 중 engine\_rpm은 speed 별로 약간 차등을 두어 설정하였다. 그 밖의 기상조건은 0으로 설정했다. 이 데이터로 테스트한 결과 만족할 만한 결과를 얻을 수 있었다. 예시 그래프와 유사한 그래프가 그려졌다.

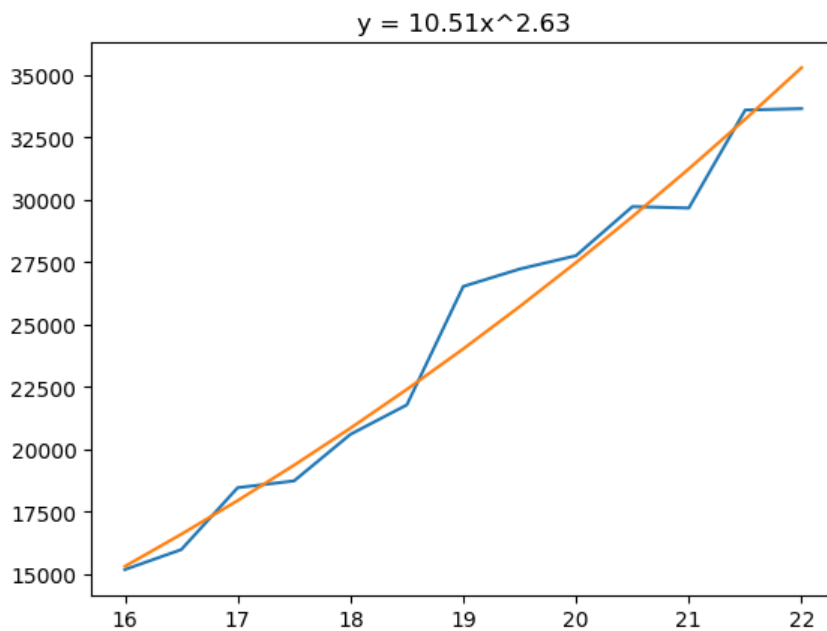


다음으로, 이 결과를 토대로 정수 중 speed - power 곡선을 추정해보기로 하였다. 추정에는 다음과 같은 방법을 사용했다.

$$y = ax^b$$

$$\log(y) = \log(a) + b\log(x)$$

Speed - power 곡선이  $y = ax^b$  를 만족한다고 가정하고 로그 식으로 바꾼 뒤, numpy의 polyfit 함수를 사용하여 데이터에 fitting된 일차함수를 구하였다. 그 계수를 통해 a, b 값들을 결정하였다. 아래의 그래프는 그 결과로 추정된 speed - power 곡선이다. 속력 V와 추진동력 P 사이에  $P \sim V^{2.63}$  가 성립함을 확인하였다. 저속에서 일반적으로  $P \sim V^{2.5}$  관계를 갖는다는 것을 감안할 때 이 결과는 실제 관계를 잘 추정하였음을 알 수 있다.



## 4.2 Conclusion

1. 이번 과제를 통해 RandomForestRegressor를 사용해봄으로써 연속형 데이터에 대한 회귀 분류를 실습해볼 수 있었다. 또한 RMSE 계산을 통하여 학습 모델에 대한 평가를 진행해볼 수 있었다.
2. 이번 과제를 통해서 전공과 좀 더 관련된 데이터를 토대로 회귀 분석을 해볼 수 있었다.