

효율적인 항만 운영 시스템을 위한 기계학습 기반 항만 내 선박 대기시간 예측

BII LAB | 학부연구생 최민화



경상국립대학교

목차

01 서론

- 1.1 문제점
- 1.2 제안하는 연구

02 관련 연구

- 2.1 관련연구

03 실험 방법

- 3.1 실험 개요 및 데이터 셋
- 3.2 전처리 및 feature selection
- 3.3 데이터 세트의 구분과 모델 성능평가

04 실험 결과

- 4.1 연구5가지 회귀 모델 성능 비교
- 4.2 하이퍼파라미터 튜닝 및 성능 평가

05 결론

- 5.1 결론

06 참고문헌

- 6.1 참고문헌

1. 서론

1.1 문제점

1.2 제안하는 연구

2년도 안 지났는데... 팬데믹 이후 최대 수출물류난

🕒 2024.06.14

아시아 환적항만들 체선 체화로 몸살

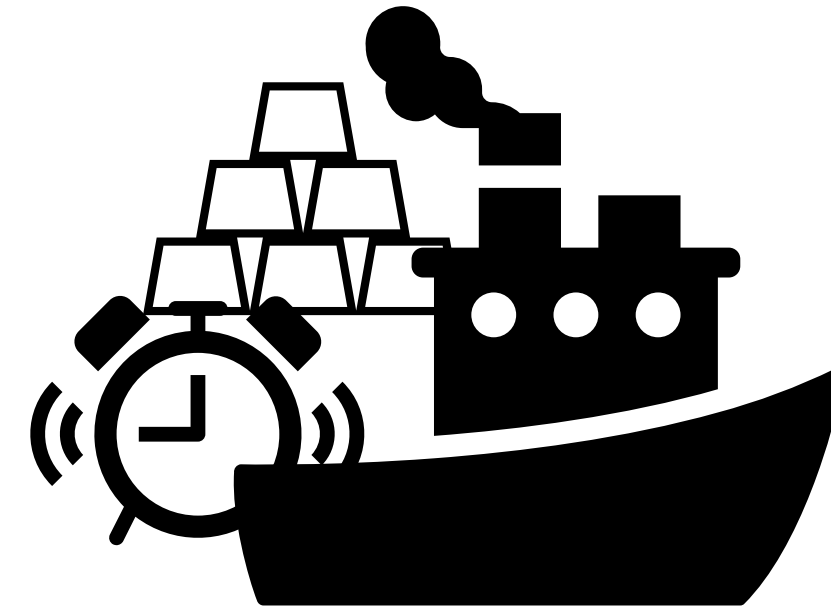
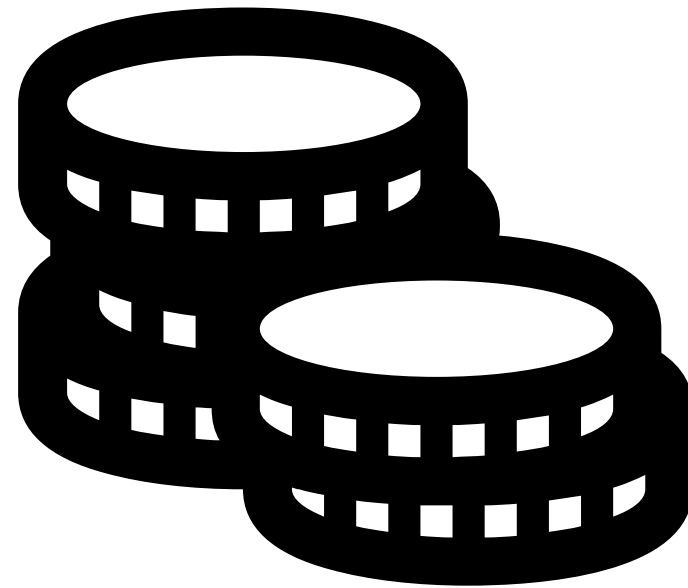
🕒 입력 2024.07.08 18:46

| 224만teu 컨테이너선 외항 대기 상태

아시아항만들의 체선·체화 현상이 점차 심각해지고 있다. 이 중에서도 세계 최대의 환적항인 싱가포르항에서는 3일-5일 정도 선박이 외항에 대기하는 체선이 계속되고 있고, 야드 내에서도 화물적체가 심화되면서 항만 오퍼레이션에 까지 악영향을 미치고 있다고 한다. 싱가포르항에서 가까운 말레이시아 '포트클랑'항에서도 비슷한 체선상태가 계속되어 대체 기항지를 찾지 못한 선사들은 아예 기항을 포기하는 사태까지 벌어지고 있는 것으로 알려졌다.

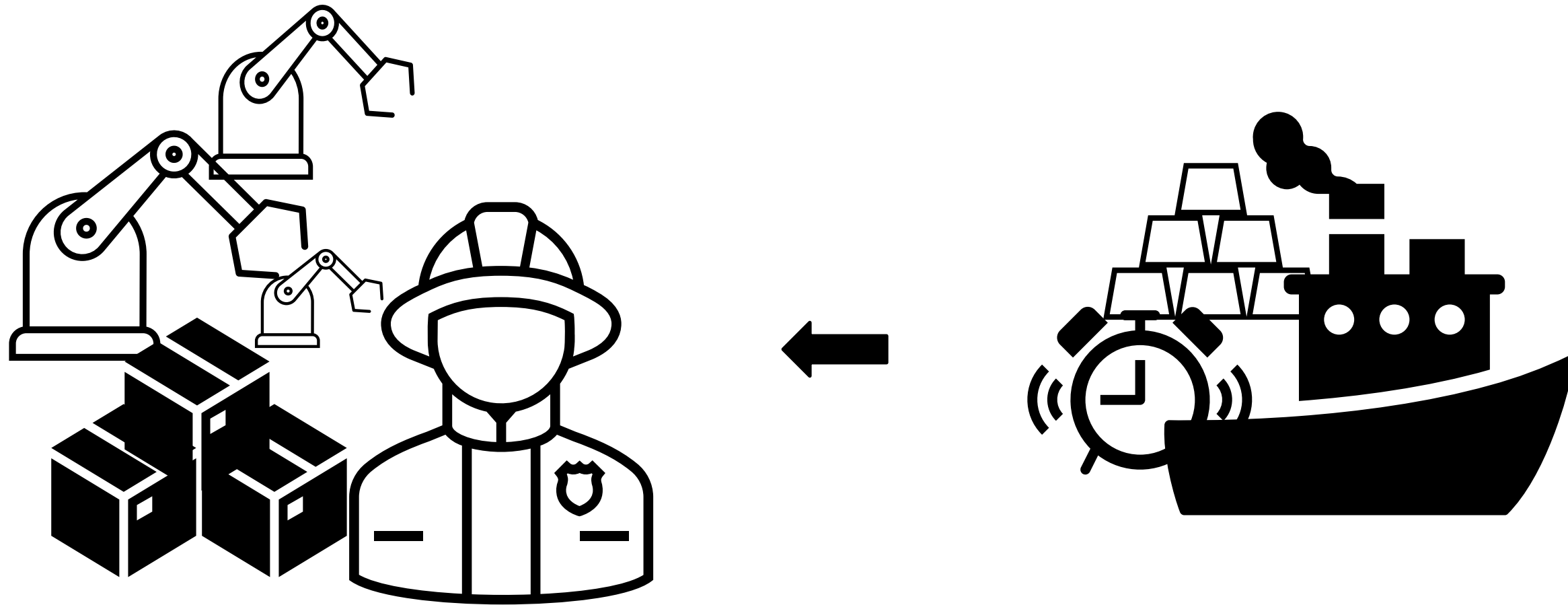


채산성 ↔ 체선현상



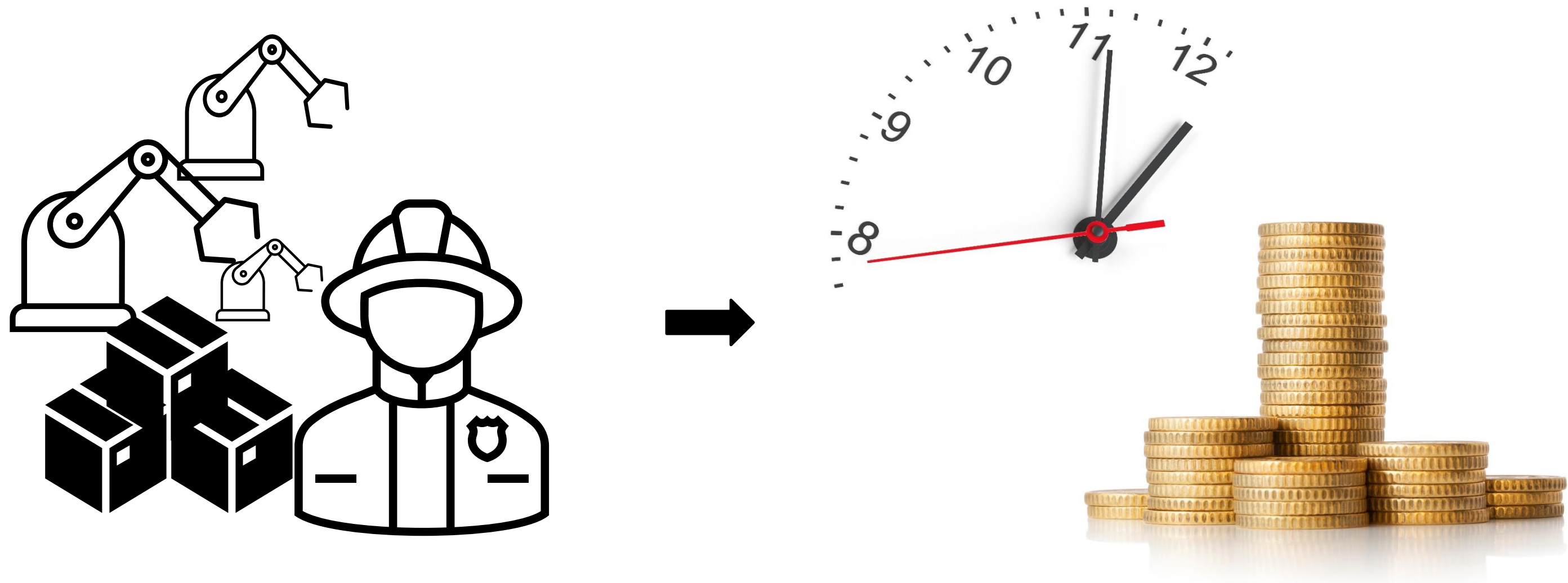
- ✓ 채산성? 수입과 지출이 맞아서 이익이 있는 성질
- ✓ 체선? 선박이 하역작업과 관련없이 항내에서 대기하는 시간

문제점 제안하는 연구



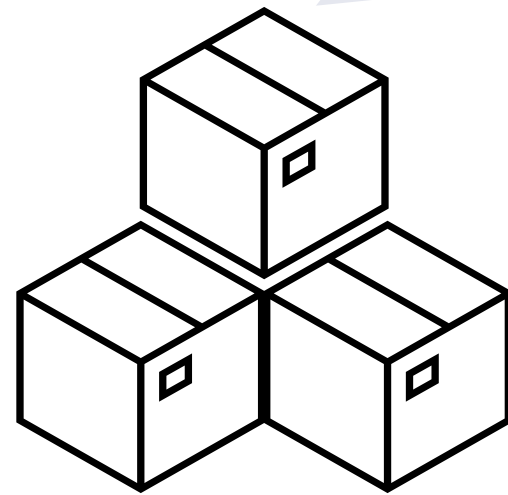
이러한 체선 현상을 해결하기 위한 항만 시설 확충은..

문제점 제안하는 연구

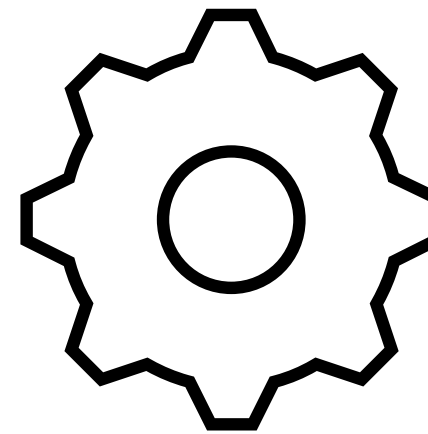


시간이 오래 걸리고 비용이 많이 듦

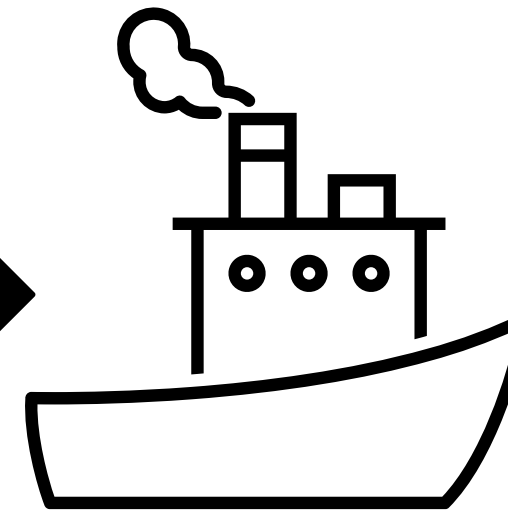
정박지와 접안지 사이의 거리, 선종, 실제 정박 시각, 도착 항의 소속 국가, 재화 중량톤수, 접안지 폴리곤 크기, 도착항의 항구명



7가지 관련 요인



예측 모델



항만 내 선박 대기시간



항만 내 선박의 대기시간을 예측하는 알고리즘 제안
항만 내 체선 현상을 더 효율적으로 개선하는 데 도움이 되고자 함

2. 관련 연구

2.1 관련 연구

관련 연구

관련 연구

① 반출입 차량의 평균 대기시간 예측

- 부산 신항의 컨테이너 반출입 정보를 기반으로 RNN, LSTM 및 GRU를 활용해 반출입 차량의 평균 대기시간을 예측
 - 부산 신항의 실제 컨테이너 반출입 데이터를 활용해 RNN, LSTM, GRU 모델의 성능 검증
- 반출입 차량의 대기시간 예측에 효과적임을 보임

② 웨이퍼 로트의 대기시간 예측

- 웨이퍼 로트의 대기시간을 예측하기 위해 심층학습 기반의 예측 모델 제안
- 공정 시작 시점의 제공 정보(work-in-process)와 디스패칭 룰 (dispatching rule)을 기반으로 대기시간에 영향을 미치는 예측 변수를 최적화함
- 유전 알고리즘을 통해 최적의 조합 도출

[4]Y. I. Kim, "A Study on the Prediction of Gate In-Out Truck Waiting Time in the Container Termina", J . Navig. P ort Res. vol. 46, no.4, pp. 344-350, August 2020.

[5]S. C. Park, "Queue-Time Prediction for a Wafer Lot with Time-Constraints", Korean Journal of Computational Design and Engineering, vol. 25, no. 4, pp. 343-349, December 2020.

3. 실험 방법

- 3.1 실험 개요 및 데이터 셋
- 3.2 전처리 및 feature selection
- 3.3 데이터 세트의 구분과 모델 성능평가

실험 개요 및 데이터 셋

전처리 및 feature selection

데이터 세트의 구분과 모델 성능평가

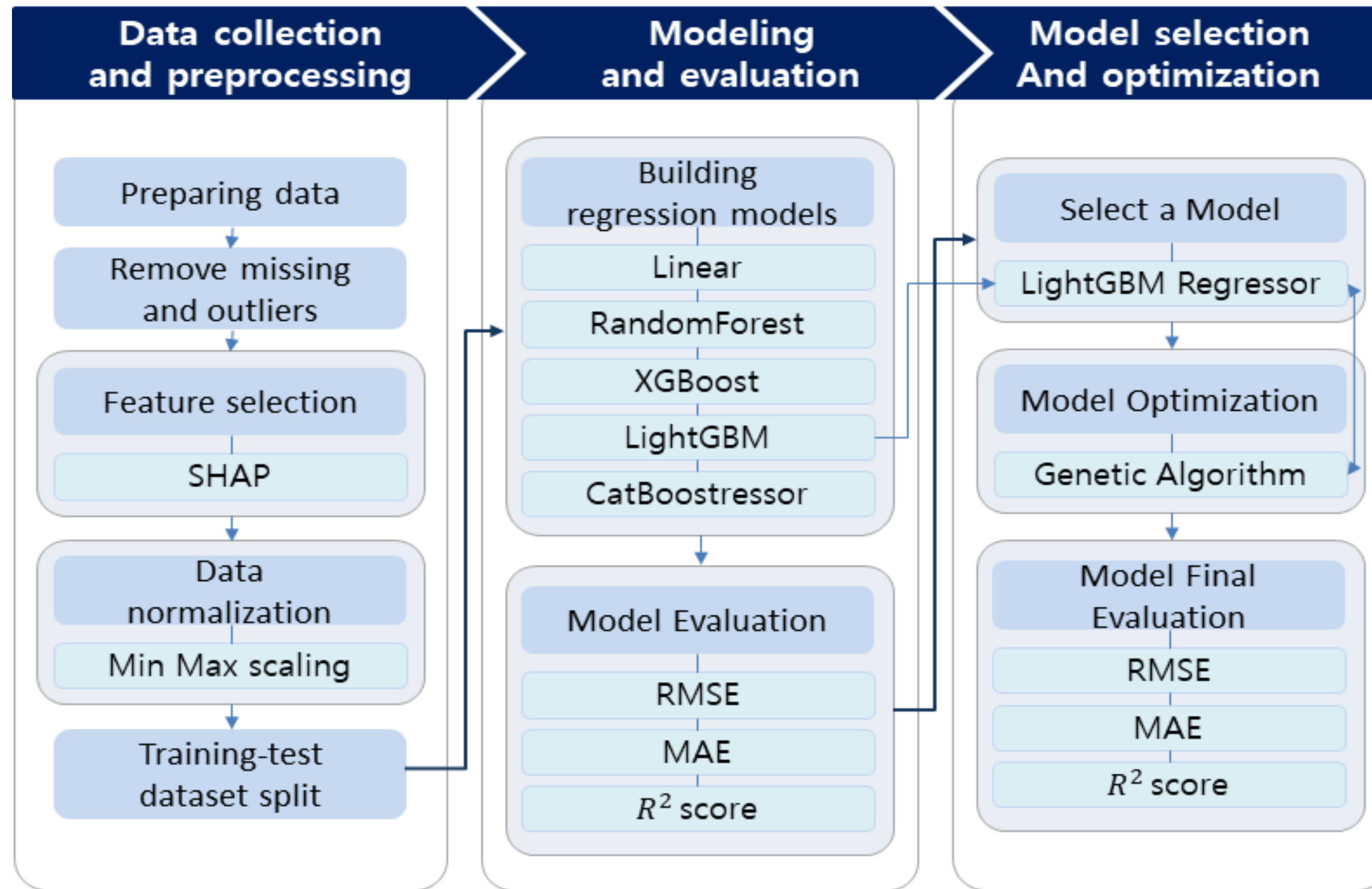


그림 1. 항구 내 선박 대기시간 예측 알고리즘 흐름도

Fig. 1. Flowchart of the algorithm for predicting ship waiting time in port

데이터 셋

- ✓ DAICON 대회 중 HD 현대 AI Challenge 데이터 셋 활용
 - 선박의 폭, 연령, 국적, 소유주, 정박지와 접안지 사이의 거리 등 총 23개의 Columns과 391,939 Rows의 개별 운행 기록에 관한 정보를 담고 있음
- ✓ 입력 데이터
 - dist (*km*) : 정박지와 접안지 사이 거리
 - ship type category : 선종
 - ata (*hour*) : 실제 정박 시각
 - ari co : 도착항의 소속 국가
 - dead weight (*ton*) : 재화중량톤수
 - port size (*km²*) : 접안지 폴리곤 크기
 - ari po : 도착항의 항구명
- ✓ 출력 데이터
 - ci hour : 대기시간

표 1. 선택된 주요 요인들의 분포

Table 1. Distribution for selected key features

Statistic	dist (km)	dead weight (ton)	port size (km ²)	ata (minute)
Min	0.000	100.000	0.000	0.000
Max	199.980	404000.000	0.002	1439.000
Mean	19.328	63577.506	0.001	689.169
Median	6.929	47400.000	0.001	659.000
Std	30.930	63515.531	0.001	421.338

전처리

- ✓ 164,631개의 결측치와 중복 항목 등을 제거
- ✓ 인코딩 통해서 범주형 자료를 연속형 자료로 변환
- ✓ 대기시간 열에 대하여 IQR (Inter Quantile Range)을 계산하여 이상치 정의 및 제거
- ✓ Min-max 정규화를 수행

Feature selection

- ✓ SHAP (Shapley Additive ex Planations) 기법 사용해 상위 7개 요인들을 독립변수로 선택
 - 정박지와 접안지 사이의 거리, 선종, 실제 정박 시각, 도착 항의 소속 국가, 재화중량톤수, 접안지 폴리곤 크기, 도착 항의 항구명

표 2. SHAP값 기준 상위 7개의 주요 요인들

Table 2 Top 7 key features based on SHAP values

순위	Feature names	SHAP
1	dist (정박지와 접안지 사이 거리)	14.035
2	ship_type_category (선종)	2.818
3	ata (실제 정박 시각)	1.143
4	ari_co (도착항의 소속국가)	0.882
5	dead weight (재화중량톤수)	0.849
6	port size (접안지 폴리곤 크기)	0.783
7	ari_po (도착항의 항구명)	0.728

데이터 세트 구분

- ✓ 5-fold cross validation을 사용
- ✓ Training 데이터 세트와 test 데이터 세트를 8:2의 비율로 분리
 - 전체 121,402 건
 - training : 약 97,121건, test : 약 24,281건

모델 성능 평가

- ✓ 5가지 회귀 모델 사용
 - ✓ Linear Regression (LR), Random Forest Regressor (RFR), XGBoost Regressor(XGBR), LightGBM Regressor (LGBR), CatBoost Regressor (CBR)
- ✓ 모델의 성능 평가 : RMSE, MAE, R^2 값 사용

4. 실험 결과

4.1 5가지 회귀 모델 성능 비교

4.2 하이퍼파라미터 튜닝 및 성능 평가

5가지 회귀 모델 성능 비교

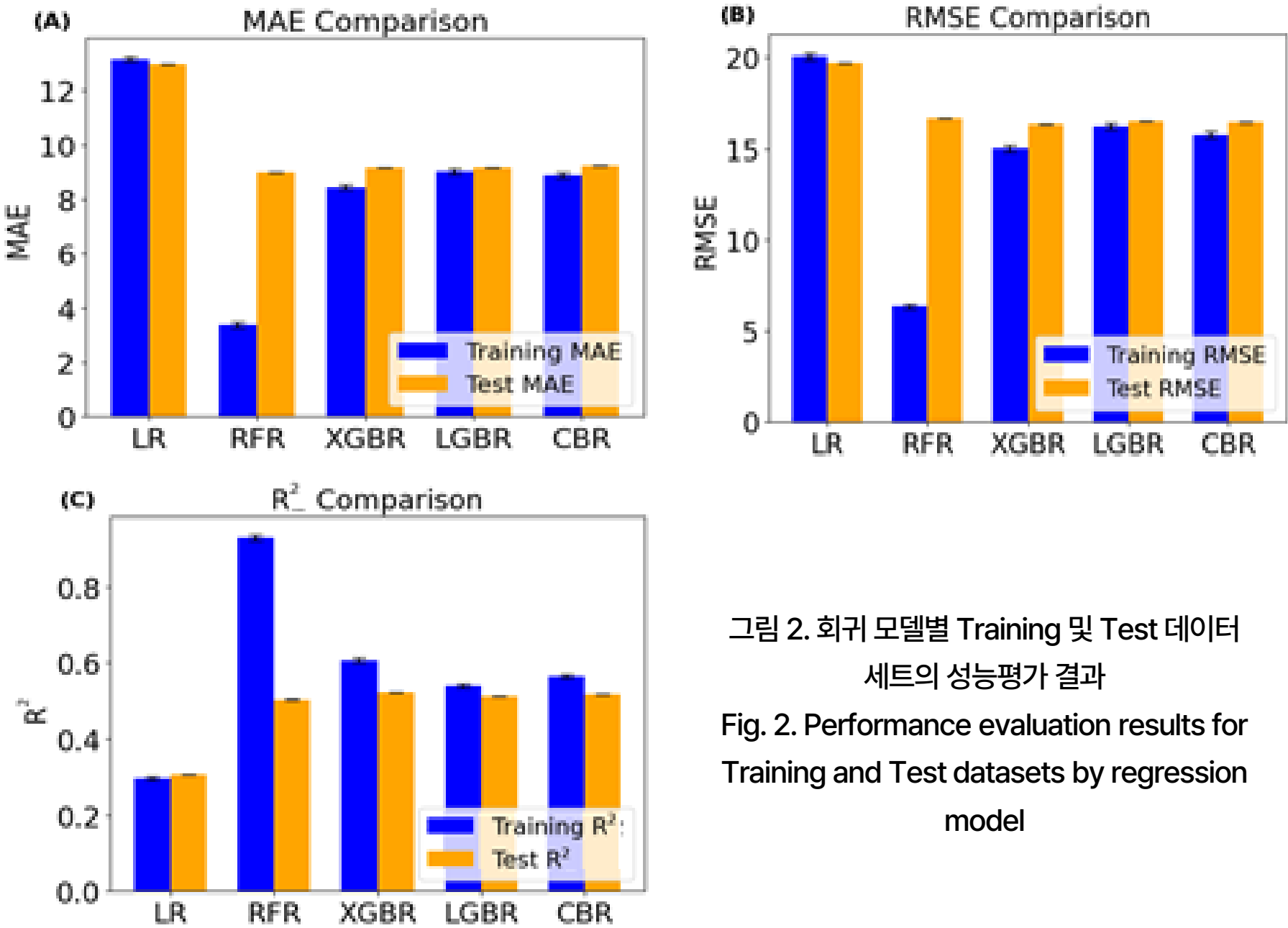
하이퍼파라미터 튜닝 및 성능 평가

5가지 회귀 모델의 Training 및 test 데이터 세트의 성능평가 비교

표 3. 회귀 모델별 Training 및 Test에서의 성능평가 결과

Table 3. Performance evaluation results for training and test by regression model

Model	Data set	RMSE	MAE	R ²
LR	Training	20.049	13.165	0.296
	Test	19.703	12.988	0.305
RFR	Training	6.319	3.395	0.930
	Test	16.637	8.984	0.504
XGBR	Training	15.012	8.444	0.605
	Test	16.321	9.178	0.523
LGBR	Training	16.213	9.031	0.540
	Test	16.489	9.163	0.513
CBR	Training	15.758	8.892	0.565
	Test	16.432	9.247	0.517



5가지 회귀 모델 성능 비교

하이퍼파라미터 튜닝 및 성능 평가

5가지 회귀 모델의 RMSE 값 분포 비교 및 분석

표 4. 회귀 모델별 test의 RMSE 평균 및 표준편차

Table 4. RMSE mean and standard deviation of test by regression model

Model	RMSE 평균	표준편차
LR	20.049	0.182
RFR	16.976	0.147
XGBR	16.619	0.182
LGBR	16.701	0.180
CBR	16.665	0.181

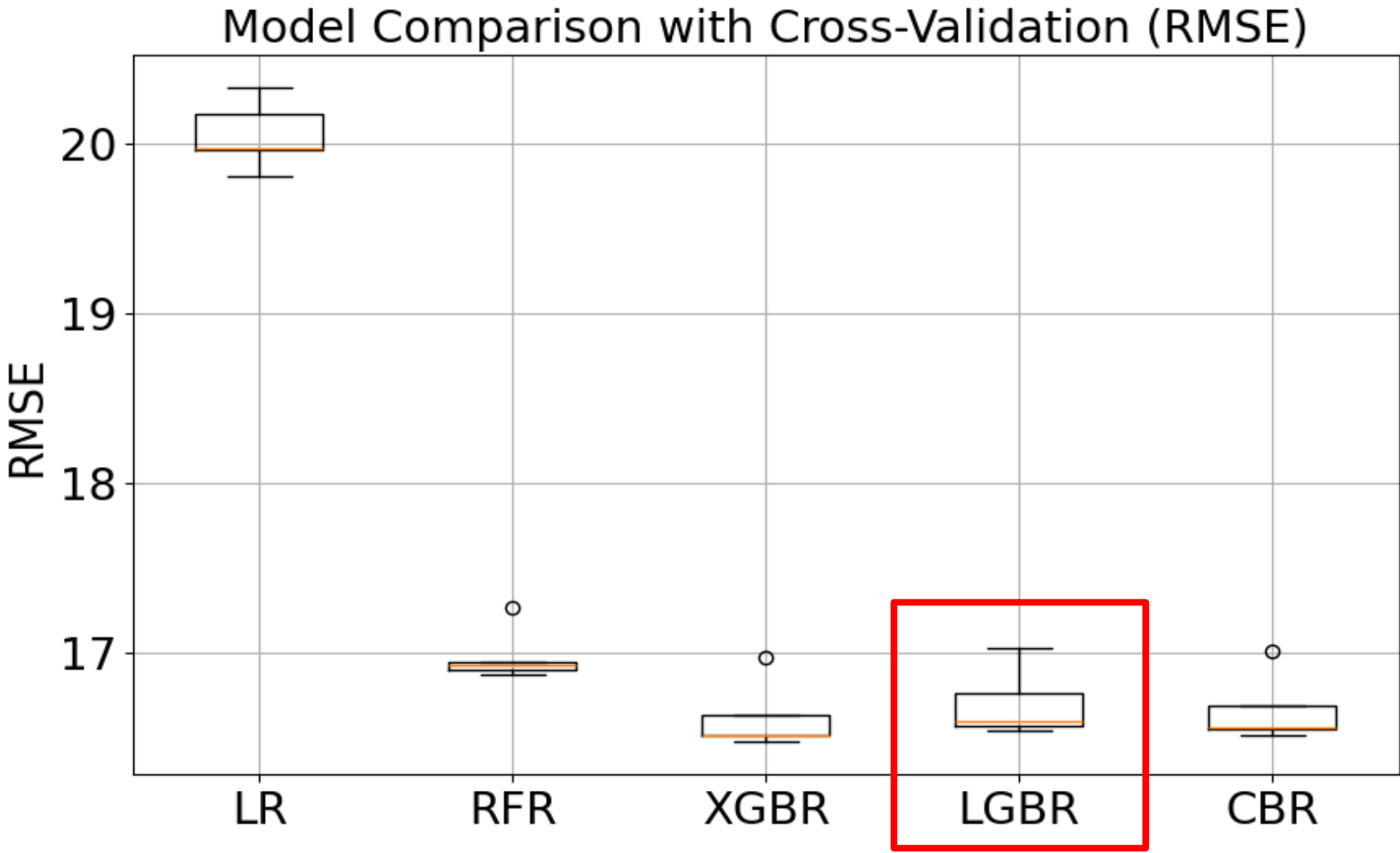


그림 3. 5가지 회귀 모델별 RMSE 결과

Fig. 3. RMSE results by five regression models

하이퍼파라미터 튜닝

- ✓ LGBR의 하이퍼파라미터 튜닝
- ✓ 유전 알고리즘 (GA, Genetic algorithm)을 사용

표 5. LightGBM Regressor의 튜닝된 하이퍼파라미터
Table 5. Tuned hyperparameters of the LightGBM Regressor

Hyper Parameter	Before	After
num_leaves	31	84.121
max_depth	-1	91
learning rate	0.1	0.072
n_estimators	100	337
min_child_weight	0.001	7.706
subsample	1.0	0.907
colsample_bytree	1.0	0.640

하이퍼파라미터 튜닝 전후 성능 비교

표 6. LightGBM Regressor 하이퍼파라미터 튜닝 전후 성능 비교

Table 6. Performance comparison of LightGBM Regressor before and after hyper-parameter tuning

Parameter Tuning	Data set	RMSE	MAE	R ²
Before	Training	16.213	9.031	0.540
	Test	16.489	9.163	0.513
After	Training	14.760	8.607	0.619
	Test	16.307	9.455	0.524

5. 결론

5.1 결론 및 향후과제

결론 및 향후과제

- ✓ 정박 대기시간은 항만의 서비스 품질을 결정하는 중요한 요소임
- ✓ 항만의 체선 현상을 개선하고자 항만 내 선박의 대기시간을 예측하는 알고리즘을 제안함
 - 정박 대기시간 예측의 정밀도 향상을 목표로 모델들을 비교한 후 LGBR 모델을 선택하였으며, 최적화 과정을 거쳐 예측 모델의 성능을 개선함
- ✓ 본 연구를 통해 예측된 선박 대기시간을 기준으로 하여 체선 현상에 미리 대비할 수 있으며, 선박 스케줄링을 효율적으로 개선 가능
- ✓ 예측 결과는 항만 관리자 뿐만 아니라 각 선박 운항자에게도 제공되어 항만의 서비스 만족도를 향상시킬 수 있을 것으로 기대됨

6. 참고 문헌

6.1 참고 문헌

참고 문헌

- [1]Kim, JH, Kim, SW, Nam, BA, Choi, JW, Jang, DW, Kim, KH and Hong, "Arrival Pattern Analysis of Gate In-Out Trucks in Container Terminal", Korean Institute of Industrial Engineers, 2015 Fall Conference of Korean Institute of Industrial Engineers, pp. 1097-1102, 2015.
- [2]Vasiliauskas A.V., Barysiene J., Dragan D., "An economic evaluation model of the logistic system based on container transportation", Transport, 23 (4), pp. 311-315, 2008.
- [3]K. H. Sa and S. B. Choi, "Empirical Study on the Congestion of Bulk Ships- Focus on the Queuing Time for Berthing", Shipping and Logistics Research vol 25, no. 2, pp. 263-284, August 2009.
- [4]Y. I. Kim, "A Study on the Prediction of Gate In-Out Truck Waiting Time in the Container Termina", J . Navig. P ort Res. vol. 46, no.4, pp. 344-350, August 2020.
- [5]S. C. Park, "Queue-Time Prediction for a Wafer Lot with Time-Constraints", Korean Journal of Computational Design and Engineering, vol. 25, no. 4, pp. 343-349, December 2020.
- [6]DACON, (2024), [train.csv], Retrieved from <https://dacon.io/competitions/official/236158/data>.

Acknowledge



윤웅창

Woongchang Yoon

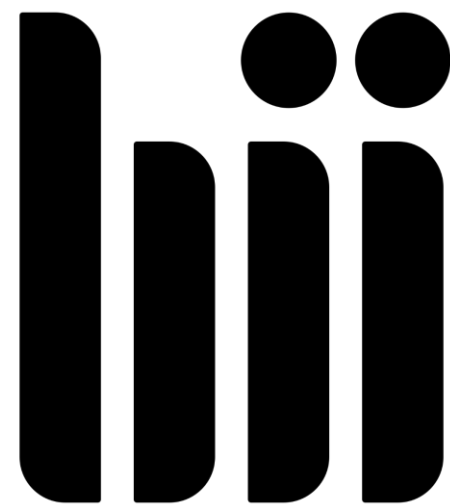
Major | bioinformatics
artificial intelligence
big data



최민화

Minhwa Choi

Major | Computer science
A field of interest | AI, ML



생명정보지능연구실
BIOINFORMATICS & INTELLIGENCE LAB

Bio Informatics Intelligence Laboratory

Department of Computer Science and Engineering, Gyeongsang National University

The field of study | Big Data Analysis

Artificial intelligence (including machine learning)

Bioinformatics