

머신러닝 분류 알고리즘을 활용한 선박 접안속도 영향요소의 중요도 분석

이형탁* · 이상원** · 조장원*** · 조익순****†

* 한국해양대학교 해양과학기술전문대학원 박사과정, ** 고베대학교 대학원 박사과정,
*** 한국해양수산연수원 교수, **** 한국해양대학교 해사글로벌학부 교수

Analysis of Feature Importance of Ship's Berthing Velocity Using Classification Algorithms of Machine Learning

Hyeong-Tak Lee* · Sang-Won Lee** · Jang-Won Cho*** · Ik-Soon Cho****†

* Ph.D. program, Ocean Science and Technology School, Korea Maritime & Ocean University, Busan 49112, Korea

** Ph.D. program, Graduate School, Kobe University, Kobe 658-0022, Japan

*** Professor, Korea Institute of Maritime and Fisheries Technology, Busan 49111, Korea

**** Professor, Division of Global Maritime Studies, Korea Maritime & Ocean University, Busan 49112, Korea

요 약 : 선박이 접안할 때 발생하는 접안에너지에 가장 영향력이 큰 요소는 접안속도이며, 과도한 경우 사고로 이어질 수 있다. 접안속도의 결정에 영향을 미치는 요소는 다양하지만 기존 연구에서는 일반적으로 선박 크기에 제한하여 분석하였다. 따라서 본 연구에서는 다양한 선박 접안속도의 영향요소를 반영하여 분석하고 그에 따른 중요도를 도출하고자 한다. 분석에 활용한 데이터는 국내 한 탱커부두의 선박 접안속도를 실측한 것을 바탕으로 하였다. 수집된 데이터를 활용하여 머신러닝 분류 알고리즘인 의사결정나무(Decision Tree), 랜덤포레스트(Random Forest), 로지스틱회귀(Logistic Regression), 퍼셉트론(Perceptron)을 비교분석하였다. 알고리즘 평가 방법으로는 혼동 행렬에 따른 모델성능 평가지표를 사용하였다. 분석 결과, 가장 성능이 좋은 알고리즘으로는 퍼셉트론이 채택되었으며 그에 따른 접안속도 영향요인의 중요도는 선박 크기(DWT), 부두 위치(Jetty No.), 재화상태(State) 순으로 나타났다. 이에 따라 선박 접안 시, 선박의 크기를 비롯하여 부두 위치, 재화 상태 등 다양한 요인을 고려하여 접안속도를 설계하여야 한다.

핵심용어 : 선박 접안속도, 머신러닝, 분류 알고리즘, 중요도, 혼동행렬

Abstract : The most important factor affecting the berthing energy generated when a ship berths is the berthing velocity. Thus, an accident may occur if the berthing velocity is extremely high. Several ship features influence the determination of the berthing velocity. However, previous studies have mostly focused on the size of the vessel. Therefore, the aim of this study is to analyze various features that influence berthing velocity and determine their respective importance. The data used in the analysis was based on the berthing velocity of a ship on a jetty in Korea. Using the collected data, machine learning classification algorithms were compared and analyzed, such as decision tree, random forest, logistic regression, and perceptron. As an algorithm evaluation method, indexes according to the confusion matrix were used. Consequently, perceptron demonstrated the best performance, and the feature importance was in the following order: DWT, jetty number, and state. Hence, when berthing a ship, the berthing velocity should be determined in consideration of various features, such as the size of the ship, position of the jetty, and loading condition of the cargo.

Key Words : Ship's berthing velocity, Machine learning, Classification algorithm, Feature importance, Confusion matrix

* First Author : gudxkr518@kmou.ac.kr

† Corresponding Author : ischo@kmou.ac.kr, 051-410-5072

1. 서론

선박이 접안할 때 접안에너지가 발생하게 된다. 접안에너지를 산정하는데 있어 가장 영향력이 큰 요소는 접안속도(Berthing Velocity)이다(Jun et al., 2008; Cho et al., 2018). 따라서 선박은 적정 접안속도로 접안하지 않으면 계류시설의 파손과 선체의 손상과 같은 사고를 유발 할 수 있다(Lee et al., 2019).

접안속도는 Brolsma et al.(1977)의 연구를 시작으로 Fig. 1 과 같이 선박 크기(Deadweight tonnage)가 증가 할수록 접안속도가 감소하는 밀접한 관계가 있다는 가정을 기초로 분석이 이루어졌다. British Standard와 Spanish ROM의 접안속도 규정에도 선박 배수량(Displacement)과의 관계를 통해 접안속도의 기준을 제시한다(PIANC, 2020). 국내 항만 및 어항설계기준에서 선박 접안속도는 재화상태, 부두 위치, 기상, 예인선 등에 근거하여 접안속도를 규정한다고 언급하고 있지만 선박 크기, 선종과의 관계에 제한된 기준을 제시하고 있다(Ministry of oceans and fisheries, 2017). Cho et al.(2018)의 경우 선박 크기, 부두 위치, 접안각도를 바탕으로 통계적 기법으로 접안속도를 분석하였지만 다른 영향요소는 고려하지 않았다. Lee et al.(2019)의 연구는 위의 요소에 선박의 재화상태만을 추가하여 확률분포 분석을 실시하였다.

하지만 접안속도를 결정하는데 영향을 미치는 요인은 다양하며 PIANC(The World Association for Waterborne Transport Infrastructure) Working Group 145에서는 이를 9가지로 분류하고 있다(PIANC, 2020). 이에 따라 접안속도에 영향을 미치는 다양한 요소를 고려하고 그 중요정도를 정량적으로 분석할 필요성이 있다. Roubos et al.(2017)는 다양한 접안속도 영향요소 데이터를 수집하고 분석하였지만 기초적인 통계 분석만을 사용하였다. 또한 기존의 가장인 선박의 크기, 재화상태, 기상 등과 접안속도의 규칙성은 찾을 수 없으며 접안각도, 도선사, 예인선 등의 영향요소가 접안속도와 상관관계가 높은 것으로 분석하였다. 하지만 기초 통계 분석의 한계로 이는 요인 간 상관관계와 중요도는 고려되지 않았다.

선박 접안속도 영향요인을 고려하고 상관관계를 도출하기 위해 본 연구에서는 머신러닝 분류 알고리즘을 활용하고자 한다. 머신러닝 기법은 복잡한 데이터에서 의미 있는 결과를 얻을 수 있는 장점이 있다(Shalev and Ben, 2014). 특히 머신러닝 분류 알고리즘은 입력한 변수들의 상관관계를 통해 중요도를 도출할 수 있다. 분석에 활용한 머신러닝 분류 알고리즘은 의사결정나무(Decision Tree), 랜덤포레스트(Random Forest), 로지스틱회귀(Logistic Regression), 퍼셉트론(Perceptron)이며, 모델의 성능을 혼동행렬(Confusion matrix)을 통한 평가 지표로 비교하고 가장 성능이 좋은 모델을 채택하여 접안속도 영향요소의 중요도를 도출하였다.

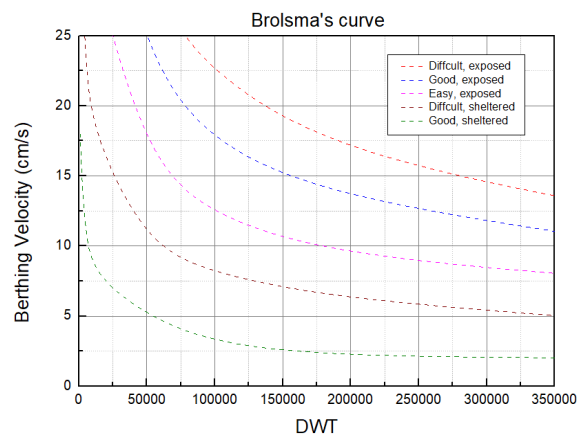


Fig. 1. Brolsma's curve.

2. 연구방법

본 연구 방법은 Fig. 2와 같은 구성으로 이루어진다.

먼저, 선박 접안관련 데이터를 수집하고 이를 바탕으로 탐색적 자료 분석을 실시한다. 머신러닝 분류 알고리즘의 적용을 위해 데이터 전처리 실시한다. 알고리즘은 비교를 통해 최적 모델을 선정하고, 모델에 따른 선박 접안속도 영향요소의 중요도를 도출한다.

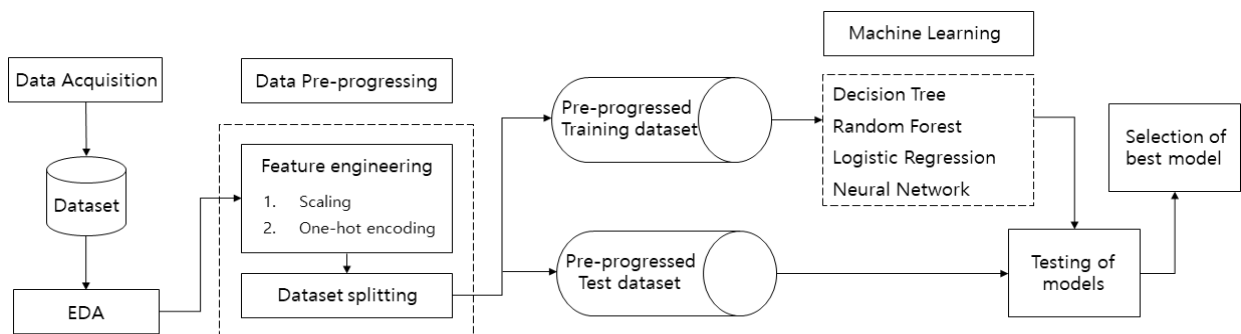


Fig. 2. Visual representation of the methodology.

2.1 데이터 수집

접안속도는 Fig. 3과 같이 전라남도 여수에 위치하고 있는 DAS(Docking Aid System)가 설치된 무역항 중 Jetty 1~3으로 구분된 tanker 터미널 부두에서 2017년 3월부터 2019년 9월까지 접안속도 실측데이터를 수집하였다. Roubos et al.(2017)은 선박이 시설물에 첫 번째 접촉할 때의 속도를 접안속도로 정의하고 있다. 따라서 Fig. 4와 같이 DAS를 통해 부두에서 선박 선체까지의 거리를 시간이 지남에 따라 측정하여 부두 시설물에 접촉한 시점의 속도를 수집하였다. 특히, ‘선박의 입항 및 출항 등에 관한 법률’에서 위험물을 취급하는 부두는 DAS와 같은 접안보조장치를 설치해야한다는 조항을 신설하여 선박의 안전한 접안을 강조하고 있다.

본 연구의 분석에 필요한 접안속도 영향요소에 해당하는 데이터 수집은 Table 1과 같이 PIANC WG145의 접안속도에 영향을 미치는 요소를 바탕으로 데이터를 수집하였다. ‘Closed/Open quay/jetty’ 요소의 경우 jetty 위치를 나타내는 변수로 지정 하였다. 접안속도를 수집한 부두는 tanker 터미널로서 선종과 화물 요인이 한정되었으므로 ‘Vessel type’의 경우, 선박의 크기에 해당하는 DWT를 수집하였으며, Lee at al.(2019)의 연구에 따라 Ballast, Half, Laden와 같이 화물적재량을 나타내었다. ‘Berthing manoeuvre’에 해당하는 접안각도는 접안속도와 같이 부두시설물에 첫 번째 접촉한 시점의 각도를 기준으로 하였다. ‘Equipment/tugs’는 선박 접안 시 배치된 tugboat의 마력의 합으로 하였으며, ‘Human factor’에 해당하는 Pilot의 경우 도선사의 경험을 나타내는 급수로 하였다. 날씨 데이터의 경우, 해당 부두에서 가장 가까운 해상부이의 기상청 데이터를 활용하였으며, Current 요인은 해당 부두가 국내의 대표적인 ‘Closed’ 항만으로 조류 영향이 거의 없는 경우에 해당하여 배제하였다. 마지막으로 ‘Berthing aids’의 경우 본 터미널 부두에서는 DAS가 설치되어 이를 활용하여 접안하고 있으며 접안속도 데이터 수집에 활용하였으므로 분석에서는 제외하였다.

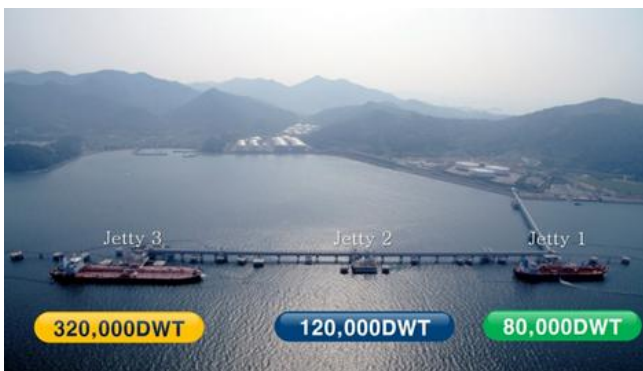


Fig. 3. Overview of tanker terminal.

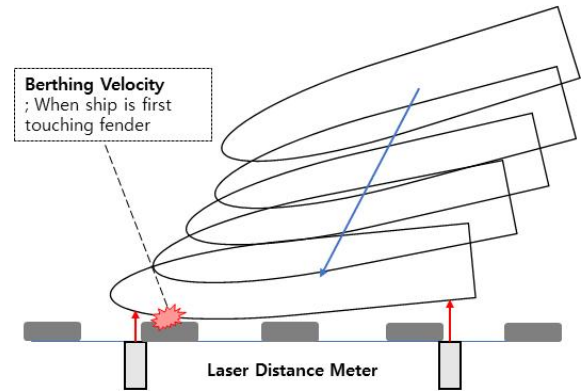


Fig. 4. Measurement of berthing velocity.

Table 1. The factor of berthing velocity in PIANC WG145

Category	Factor(WG145)	Remark
Environment	Current	Weather Condition
	Effect waves	
	Effect wind	
	Closed/Open quay /jetty	Jetty Position
	Berthing aids	Docking Aid
Vessel	Vessel type	Type of vessel
		Size of vessel
		Type of cargo
		Windage area
Human	Berthing manoeuvre	Berthing angle
	Equipment/tugs	Tugs H/P
	Human factor	Pilot

2.2 탐색적 자료 분석

탐색적 자료 분석(Exploratory Data Analysis, EDA)은 데이터의 모양을 확인하고, 기초적인 분석을 실행하는 단계이다 (Tukey, 1977). 탐색적 자료 분석은 데이터를 분석할 도구를 설정하고, 변수가 가지는 의미와 성질을 파악한 후 기초적인 분석을 하는 것으로 구성된다. 더하여, 분석을 통해 나온 결과물을 시각화하는 것도 포함된다.

2.3 데이터 전처리

데이터 전처리(Data pre-processing)는 데이터를 머신러닝 모델에 적용하기 위해 분석에 적합한 형태로 변형하는 과정이다. 이는 데이터 분석 과정에 있어 꼭 필요한 과정이며, 특히 머신러닝 모델의 성능을 높이기 위해 초기 데이터를 가공하는 과정을 거쳐야 한다(Zheng and Casari, 2018). 본 연구에서는 연속형 데이터의 단위를 통일하는 데이터 스케일링(Data Scaling)과 범주형 데이터 변환에 사용되는 원-핫 인코딩(One-hot encoding)을 사용하여 데이터를 처리하였다.

2.3.1 데이터 스케일링

데이터 스케일링은 데이터의 단위를 표준화하는 과정이다. 다차원의 데이터들을 분석할 때 데이터 단위의 차이로 인하여 분석 결과값에 오차가 생기지 않도록 표준화하여 분석해야한다(Han et al., 2011).

따라서 본 연구에서는 연속형(Continuous) 데이터의 분석을 위해 스케일링을 수행하였으며, 수행한 스케일링의 종류는 RobustScaler를 사용하였다.

RobustScaler은 식(1)과 같으며 데이터(x)에 대해 IQR (interquartile range, $Q_3 - Q_1$)과 중앙값(x_i)을 사용하여 스케일링을 하는 방식이다. 다른 스케일링 기법보다 아웃라이어 (Outlier)의 영향을 최소화한 장점이 있다.

$$\frac{x_i - Q_1(x)}{Q_3(x) - Q_1(x)} \quad (1)$$

2.3.2 원-핫 인코딩

원-핫 인코딩은 가변수라고 하며, Fig. 5와 같이 범주형 (Categorical) 변수를 이진 벡터로 표시하는 방법이다. 모든 범주형 변수를 정수인 1과 0의 이진형 벡터로 표시하며 범주형 변수를 열거하여 해당하지 않은 모든 항목은 0으로 해당하는 항목은 1로 표시하는 것이다(Harris and Harris, 2007).

또한 본 연구에서 수행해야하는 머신러닝 이진 분류 알고리즘을 위해서는 종속 변수인 접안속도를 2개의 범주로 분리해야한다. 이를 위해 데이터를 수집한 탱커부두의 설계기준 규정을 참조하여 분류하였다.

Table 2에 따르면 접안속도 10 cm/s까지를 경고속도로 규정하고 있다. 또한 터미널에서는 critical에 해당하는 속도로 접안하여 부두시설물에 손상을 가하는 경우, 해당 pilot을 black-list 처리한다는 규정을 지정하였다. 특히, 항만 및 어항 설계기준에서도 선박이 접안속도를 10 cm/s 이내로 하여 접안할 것을 권고하고 있다(Ministry of oceans and fisheries, 2017). 따라서 이를 반영하여 본 연구에서는 접안속도 10 cm/s를 기준으로 10 cm/s 이상과 10 cm/s 미만으로 나누어 분석을 진행하고자 한다.

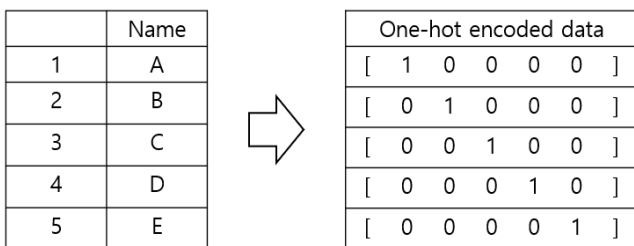


Fig. 5. One-hot encoding.

Table 2. Tanker terminal regulation of berthing velocity

Approaching velocity	Description
5 cm/sec or less	allowable velocity
6 ~ 10 cm/sec	warning velocity
Over 11 cm/sec	critical velocity

2.3.3 데이터셋 분류

홀드아웃(Hold-out method)은 Fig. 6와 같이 교차검증(Cross Validation)의 방법 중 하나로 데이터셋을 Train 데이터와 Test 데이터로 나누어 모델을 검증하는 방법이다(Kohavi, 1995).

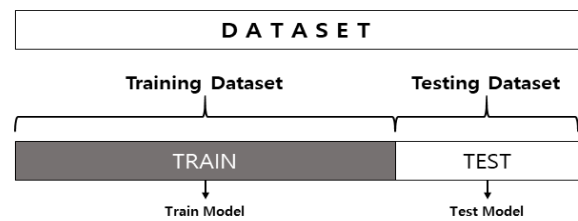


Fig. 6. Hold-out method.

따라서 본 연구에서는 머신러닝 알고리즘 구축 시 일반적으로 사용하는 홀드아웃을 적용하여 수집한 데이터를 Train 데이터셋과 Test 데이터셋으로 나누어 분석을 진행하였다.

2.4 머신러닝 분류 알고리즘

본 연구에서는 접안속도를 10 cm/s 이상과 10 cm/s 미만으로 나누어 분석하고 영향요소의 중요도를 도출한다. 따라서 목적변수가 두개로 분류된 방법을 사용해야하므로 머신러닝 이진분류 알고리즘을 활용하였다. 머신러닝 이진분류 알고리즘의 대표적인 방법은 의사결정나무, 랜덤포레스트, 로지스틱 회귀, 퍼셉트론이 있다.

2.4.1 의사결정나무

의사결정나무는 입력된 값에 대하여 예측, 분류하기 위해 노드 분할과 가지치기로 이루어진 알고리즘이다(Hastie et al., 2009). 노드 분할 기준에 따라 다양한 알고리즘이 있으며 일반적인 기법 중 하나는 CART(Classification And Regression Tree)이다(Breiman et al., 1984). 이 알고리즘은 노드에서 가지를 칠 때 이진분류(Binary Split)를 하는 방식이다. 본 연구에서 사용되는 나무의 노드 분할 척도로는 Gini Index를 사용하였다. Gini Index는 전체 데이터(m)을 임의의 D 영역을 기준으로 분리할 때 ' i ' 범주에 속하는 class의 비율(p_i)과 속하지 않는 비율($1-p_i$)의 곱으로 나타내며, 데이터가 잘못 분류되는 빈도를 측정하는 것으로 식(2)와 같다(Han et al., 2011).

$$Gini(D) = \sum_{i=1}^m p_i(1-p_i) = 1 - \sum_{i=1}^m p_i^2 \quad (2)$$

$$y = \begin{cases} 1 & \text{if } (\sum_{i=1}^N w_i x_i) \geq \theta \\ 0 & \text{if } (\sum_{i=1}^N w_i x_i) < \theta \end{cases} \quad (6)$$

2.4.2 랜덤포레스트

랜덤포레스트는 다수의 의사결정나무가 모여 숲을 이룬 형태로 모델을 학습하는 방법이다(Breiman, 2001). 숲을 구성하는 방법은 배깅(Bootstrap Aggregating, Bagging)을 기초로 한다. 배깅은 부스트스트랩(Bootstrap)을 통해 조금씩 다른 Train 데이터셋을 여러 개 생성하여 훈련하고 결합시키는 방법이다. 따라서 이 알고리즘은 $T_1(x), T_2(x), \dots, T_b(x)$ 와 같이 b 개($b = 1$ to B)로 나누어진 각각의 의사결정나무($\{T_b\}_1^B$)를 랜덤하게 생성한다. 배깅으로 생성된 여러 나무들의 예측값인 $\hat{C}_b(x)$ 중 가장 많은 선택(vote)을 받은 class로 분류하는 앙상블(ensemble) 학습이 랜덤포레스트(rf)이다. 식은 (3)과 같다.

$$\hat{C}_{rf}^B(x) = \text{majority vote} \{ \hat{C}_b(x) \}_1^B \quad (3)$$

2.4.3 로지스틱 회귀

로지스틱회귀분석은 시그모이드(sigmoid)함수를 기초로 구성된다(Shalev-Shwartz and Ben-David, 2014). 선형회귀와 달리 0과 1로 분류되므로 이진분류 알고리즘으로 유용하게 사용된다. 시그모이드 함수는 식(4)와 같다. 로지스틱 회귀의 접근방식은 독립변수(X_1, \dots, X_i)와 회귀계수(B_0, \dots, B_i)에 관한 선형 예측함수를 기초로 한다. 이에 따라 독립변수(X_1^i)가 주어졌을 때 종속변수(Y)가 1의 범주에 속할 확률을 계산한 로지스틱회귀는 식(5)와 같다.

$$\Phi_{sig}(z) = \frac{1}{1 + \exp(-z)} \quad (4)$$

$$P(Y=1|X_1^i) = \frac{1}{1 + \exp(B_0 + B_1 X_1 + \dots + B_i X_i)} \quad (5)$$

2.4.4 퍼셉트론

퍼셉트론은 Rosenblatt(1962)의 연구에서 제안된 인공신경망의 기초적인 알고리즘이다. 퍼셉트론은 다수의 신호를 입력으로 받아 하나의 신호로 출력하는 방법이다(Kim et al., 2019). 머신러닝에서는 퍼셉트론보다 더욱 발전된 Multi-Layer Perceptron(MLP)이 많이 사용되고 있으나, 본 연구의 목적인 변수별 중요도를 파악하기 위해, 가중치를 통해 시각적으로 즉시 분별할 수 있는 퍼셉트론 알고리즘을 사용하였다. 퍼셉트론은 입력변수 x_1, x_2, \dots, x_i 에 가중치 w_1, w_2, \dots, w_i 가 적용되고 그 합은 threshold θ 의 비교로 계산되며 식(6)과 같다(Kanal, 2003).

2.5 알고리즘 평가

머신러닝 분류 알고리즘의 평가는 혼동 행렬(Confusion matrix)을 기반으로 구성된다(Han et al., 2011). 종속 변수에서 구하고자 하는 범주를 긍정(Positive)이라고 하고 그 반대를 부정(Negative)라고 한다. 본 연구에서의 긍정의 범주는 10 cm/s 이상, 부정의 범주는 10 cm/s 미만이라고 할 수 있다. 혼동행렬의 구성은 Fig. 7과 같으며 그에 따른 평가방법은 Table 3과 같다.

- True Positives(TP): 실제 긍정적인 범주를 긍정이라고 예측
- True Negative(TN): 실제 부정인 범주를 부정이라고 예측
- False Positives(FP): 실제 부정인 범주를 긍정이라고 예측
- False negative(FN): 실제 긍정적인 범주를 부정이라고 예측

		Predicted class		
		yes	no	Total
Actual class	yes	TP	FN	P
	no	FP	TN	N
	Total	P'	N'	P+N

Fig. 7. Confusion Matrix.

Table 3. Evaluation measures

Measure	Formula
Accuracy	$\frac{TP + TN}{P + N}$
Error rate	$\frac{FP + FN}{P + N}$
Recall	$\frac{TP}{P}$
Specificity	$\frac{TN}{N}$
Precision	$\frac{TP}{TP + FP}$
F-1 score	$\frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$
F_β , where	$\frac{(1 + \beta^2) \times \text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}}$

3. 연구결과

3.1 기초 데이터 분석

본 연구에서는 Table 4와 같이 변수를 구분하고, 종속변수는 접안속도, 독립변수는 접안속도 영향요인으로 정의한다. 분석에 사용되는 독립변수는 3개의 범주형 변수와 5개의 연속형 변수로 이루어져 있다. 특히, 종속변수인 접안속도의 경우 머신러닝 분류 알고리즘을 적용하여 분석하기 위해서는 범주를 구분할 필요가 있어 Table 2에 따라 10 cm/s 이상(Over10)과 10 cm/s 미만(Under10)으로 나누었다.

범주형, 연속형 변수의 상세사항은 Table 6, Table 7과 같다. Jetty No., MaxAngle, Berthing Velocity는 DAS를 통해 수집하였다. State, DWT 변수는 Port-Mis를 통해 조사하였다. Pilot, Tug H.P. 변수는 도선예보를 통해 수집하였으며, Wind, Wave 변수는 기상자료개방포털에서 터미널과 가장 가까운 해상 부이의 데이터를 참조하였다. 범주형 변수는 Fig. 8에서 Frequency 그래프로 시각화하였다. 3개의 범주형 변수에서 Jetty2, Ballast, Class1의 경우 약 31.8%, 36.4%, 18.2%로 Over10가 차지하는 비중이 상대적으로 높았으나 접안속도와 명확한 규칙성은 나타나지 않았다. 연속형 독립 변수의 경우 Fig. 8과 같이 Scatter 그래프로 시각화 하였다. DWT와 Tug H.P.의 경우 R^2 가 -0.112, -0.116으로 음의 상관관계를 보였지만, R^2 값이 낮아 설명력이 떨어진다. 나머지 변수와 접안속도와의 R^2 는 0에 가까운 값을 보였다. 따라서 통계적 기법에 의한 결과로는 접안속도와 영향요소 간 관계를 설명할 수 없으므로 머신러닝 분류 알고리즘의 적용이 필요하다.

Table 4. Variable description table

Name	Type	Description	Units
Jetty No.	Categorical	Jetty Number (1~3)	-
State	Categorical	Ballast, Half, Laden	-
Pilot	Categorical	Pilot Class (1st~4th)	-
DWT	Continuous	Deadweight tonnages	tons
MaxAngle	Continuous	Max berthing angle	deg
Tug H.P.	Continuous	Tugs horse power	HP
Wind	Continuous	Max wind speed	m/s
Wave	Continuous	Significant wave height	m
Berthing Velocity	Categorical	Over 10 cm/s, Under 10 cm/s	-
	Continuous	Ship's berthing velocity	cm/s

3.2 머신러닝 모델 구축 및 평가

머신러닝 모델 구축 단계에 앞서, 데이터 스케일링과 원-핫 인코딩을 통해 데이터의 전처리를 실시하였다. 전처리를 완료한 데이터는 홀드아웃 방식을 사용하여 Table 5와 같이 3:1비율로 Train 데이터와 Test 데이터를 분리하였다. 머신러닝 모델을 학습하기 위한 Train 데이터는 267개, 모델의 성능을 평가하는 Test 데이터는 89개로 구성되어 있다.

Table 5. Description of train and test dataset

	Dependent	Count	Total
Train data	Over10	50	267
	Under10	217	
Test data	Over10	22	89
	Under10	67	

Train 데이터를 활용하여 의사결정나무, 랜덤포레스트, 로지스틱회귀, 퍼셉트론을 모델을 구축하였다. 구축한 모델에 Test 데이터를 입력한 결과는 혼동행렬로 나타나며 Table 8과 같다. 혼동행렬을 바탕으로 평가방법을 적용한 결과는 Table 9와 같다. 본 연구에서는 종속변수가 접안속도 10cm/s를 기준하므로, 긍정의 범주에 해당하는 Over10의 분류정도를 나타내는 Recall과 Precision지표를 활용하였다. Recall은 실제 입력된 22개의 Over10값 중 Over10으로 올바르게 예측된 지표이다. Precision은 Over10으로 예측한 데이터 중 실제 Over10값으로 정확히 분류한 지표이다. Recall과 Precision을 활용하여 모델을 평가하는 방법은 F-1 score이며, 지표에 가중치를 적용하여 평가하는 방법이 F_β 이다. 따라서 평가지표는 F-1 score와 F_β 값을 활용하였는데 특히 F_β 는 예측의 정확성을 나타내는 Precision에 가중치를 두었다. 모델평가 결과, F-1 score와 F_β 지표는 퍼셉트론, 의사결정나무, 랜덤포레스트, 로지스틱 회귀 순으로 나타났으며 그에 따라 퍼셉트론(F-1 score:0.476, F_β :0.490)을 최종 모델로 채택하였다.

3.3 접안속도 영향요소 중요도 분석

머신러닝 분류 알고리즘에 독립변수로 입력된 접안속도 영향요소는 분류 순서, 변수에 대한 가중치 등으로 변수의 중요도를 분석할 수 있다. 이는 단순한 통계적 분석기법과 다르게 독립변수간의 영향을 통해 규칙성을 얻고 종속변수로 결과를 나타내므로 각 요인별 중요도를 비교하는데 적절하다고 할 수 있다. 최종 모델로 선택된 퍼셉트론의 경우, Fig. 9와 같이 각 변수 가중치의 절대값을 통해 변수 간 중요도를 도출한다(Diersen et al., 2011). 즉, 해당 변수의 가중치의 크기에 따라 1(Over10)과 0(Under10)으로 결과값을 도출하여 접안속도를 이진 분류하는데 있어 영향을 미치는 정도를 나타내는 지표이다.

머신러닝 분류 알고리즘을 활용한 선박 접안속도 영향요소의 중요도 분석

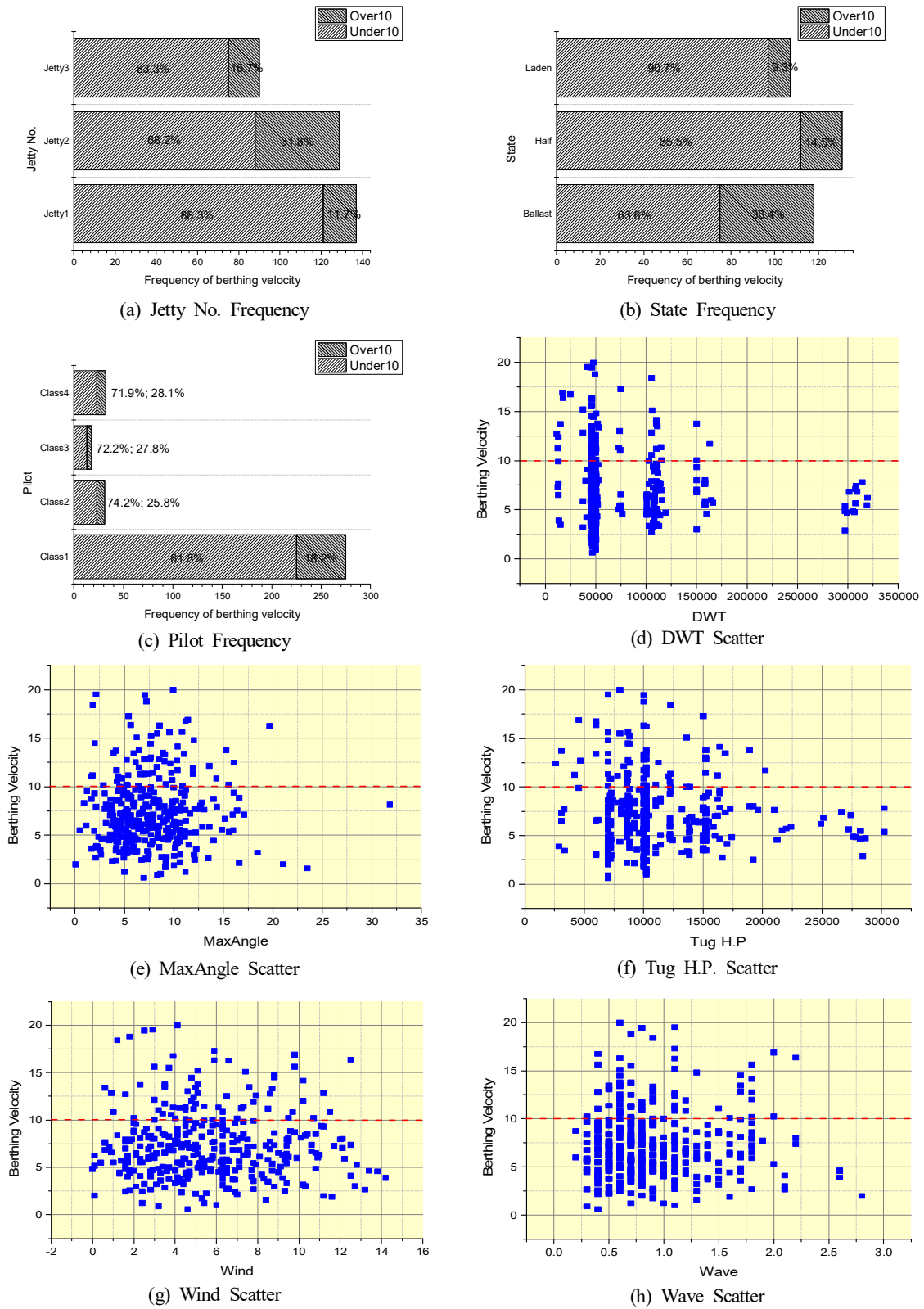


Fig. 8. Graphs between variables according to data analysis visualization.

Table 6. Descriptive statistics for categorical dataset

Berthing Velocity	Jetty No.			State			Pilot				Total
	Jetty1	Jetty2	Jetty3	Ballast	Half	Laden	Class1	Class2	Class3	Class4	
Over10	16	41	15	43	19	10	50	8	5	9	72
Under10	121	88	75	75	112	97	225	23	13	23	284
Total	137	129	90	118	131	107	275	31	18	32	356

Table 7. Descriptive statistics for continuous dataset

	DWT	MaxAngle	Tug H.P.	Wind	Wave	Berthing Velocity
Mean	73832	7.6	11144	5.7	0.9	7.4
σ	57239	3.8	4711	3.0	0.5	3.6
Min	11410	0.1	2600	0.0	0.2	0.6
25%	46176	5.1	8600	3.4	0.6	4.8
50%	49990	7.1	10000	5.2	0.8	6.9
75%	100927	9.7	13315	7.5	1.1	9.3
Max	319428	31.8	30218	14.2	2.8	20.0

Table 8. Confusion matrix of test dataset

Actual \ Predicted	Decision Tree		Random Forest		Logistic Regression		Perceptron	
	Over10	Under10	Over10	Under10	Over10	Under10	Over10	Under10
Over10	6	16	3	19	1	21	10	12
Under10	6	61	1	66	1	66	10	57

Table 9. Evaluation of machine learning classification algorithm

	Decision Tree	Random Forest	Logistic Regression	Perceptron
Accuracy	0.753	0.775	0.753	0.753
Error rate	0.247	0.225	0.247	0.247
Recall	0.273	0.136	0.046	0.455
Specificity	0.910	0.985	0.985	0.851
Precision	0.500	0.750	0.500	0.500
F-1 score	0.353	0.230	0.084	0.476
F_{β} ($\beta=0.5$)	0.429	0.394	0.168	0.490

퍼셉트론 모델의 가중치 결과를 나타낸 Fig. 10에 따르면 접안속도 영향요소 중요도는 선박의 크기(DWT, 9.5861), Jetty의 위치(Jetty No., 6.77959), 화물적재상태(State, 5.22926), Wave(2.71922), Tug Horse Power(Tug H.P., 2.07783), 도선사 급수(Pilot, 0.55061), Wind(0.34899), 최대접안각도(MaxAngle, 0.18789) 순으로 나타났다. 따라서 선박 접안 시 선박 크기, 부두 위치, 재화상태 순으로 접안속도 영향요인을 고려해야 한다. 특히 본 연구에서 접안속도를 수집한 탱커터미널의

경우 15만DWT 이하의 선박이 접안하는 경우 Over 10 cm/s에 해당하는 경우가 발생하였으므로 선박 접안 시 주의해야한다. Jetty의 위치의 경우 12만 DWT의 capacity를 보이는 Jetty 2에 접안하는 경우, 비교적 접안속도가 높은 선박의 접안이 이루어진 것을 알 수 있다. 또한 화물적재상태가 Ballast 상태일 때 상대적으로 위험 접안속도로 접안하는 경우가 있는 것으로 도출되어 주의하여 접안하여야 한다.

머신러닝 분류 알고리즘을 활용한 선박 접안속도 영향요소의 중요도 분석

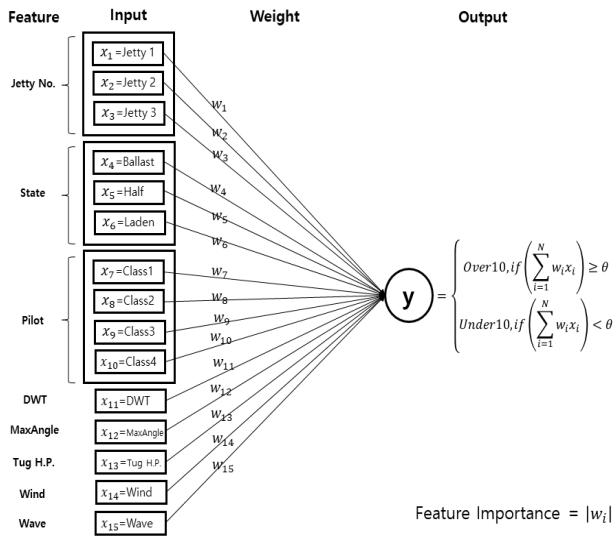


Fig. 9. Result of perceptron model.

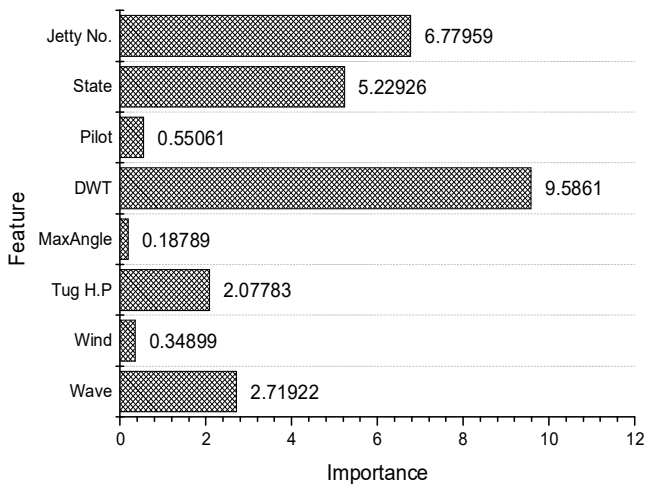


Fig. 10. Feature importance of ship's berthing velocity.

4. 결론

본 연구에서는 선박 접안속도 영향요소의 중요도를 도출하였다. 분석을 위해 국내 한 Tanker 부두에서 356개의 접안속도 데이터를 실측하고 영향요소에 대한 데이터도 함께 수집하였다. 알고리즘은 데이터를 수집한 8개의 영향요소를 고려하여 분석할 수 있는 머신러닝 분류 알고리즘을 적용하였다. 영향요소의 중요도는 혼동행렬을 사용하여 그에 따른 평가지표를 활용하였다. 그 결과를 요약하면 다음과 같다.

(1) 기존 접안속도 관련 연구에서는 선박의 크기, 부두의 위치, 재화상태 등이 접안속도에 영향을 미치는 요소라 가정된 분석이 이루어졌다. 하지만 본 연구에서는 8개의 접안

속도 영향요소를 고려하여 분석하였다. Jetty No., State, Pilot 요소에서 Over10의 비중이 높은 변수에 해당하는 Jetty No. 2, Ballast, Class1에서 32%, 36%, 18%로 나타났고, DWT와 Tug H.P.의 경우 R^2 가 음의 상관관계인 -0.112, -0.116으로 분석되었지만 미세하였으며, 나머지 요소는 R^2 가 0에 가까웠다. 따라서 기초적인 통계 분석만으로는 접안속도와 영향요소 간의 규칙성을 도출할 수 없다는 한계점이 있었다.

(2) 접안속도 영향요인을 독립변수로 설정하고 선박 접안속도를 종속변수로 설정하였다. 분석을 위해 데이터 표준화와 범주형 변수의 변환을 위해 전처리를 실시하였다. 전처리가 완료된 접안속도 데이터를 머신러닝 분류 알고리즘으로 적용하였으며 의사결정나무, 랜덤포레스트, 로지스틱회귀, 퍼셉트론을 활용하였다.

(3) 머신러닝 분류 알고리즘의 적용 결과는 혼동행렬로 나타나며 이에 따른 모델 평가지표는 F1-score, F_β 값을 활용하였다. 그 결과 퍼셉트론이 0.476, 0.49 값으로 가장 성능이 좋은 모델로 채택되었다.

(4) 퍼셉트론에 따른 접안속도 영향요소의 중요도는 가중치로 계산되며, 연구 결과 DWT, Jetty No., State, Wave, Tug H.P., Pilot, Wind, MaxAngle 순으로 나타났다. 따라서 선박이 안전하게 부두에 접안하기 위해서는 위와 같은 영향요소의 우선순위에 따라 접안속도를 고려해야 한다. 또한 DWT가 커질수록 접안속도는 낮아지며, 부두의 위치에 따라 접안속도 결정에 영향을 미치고, 화물을 적재한 경우보다 경하상태일 때 접안속도가 높은 경향을 보인다는 결과를 도출하였다.

하지만 본 연구는 국내의 한 Tanker 부두에 한정된 연구 결과라는 한계점이 있다. 더하여 데이터의 수를 보다 많이 수집할 수 있다면 분석결과의 설명력은 높아질 것이라고 판단된다. 따라서 다양한 선종과 부두를 대상으로 많은 수의 데이터를 수집하여 분석한다면 신뢰성 높은 결과를 도출할 것으로 생각한다.

References

- [1] Breiman, L.(2001), Random forests, Machine learning, Vol. 45, No. 1, pp. 5-32.
- [2] Breiman, L., J. H. Friedman, R. Olshen, and C. J. Stone (1984), Classification and Regression Trees, Wordsworth.
- [3] Brolsma, J. U.(1977), On Fender Design and Berthing Velocities, Proc. International Navigation Congress, Section II, Subject 4, pp. 87-100.
- [4] Cho, I. S., J. W. Cho, and S. W. Lee(2018), A Basic Study on the Measured Data Analysis of Berthing Velocity of

- Ships, Journal of Coastal Disaster Prevention, Vol. 5, No. 2, pp. 61-71.
- [5] Diersen, S., E. J. Lee, D. Spears, P. Chen, and L. Wang (2011), Classification of seismic windows using artificial neural networks, *Procedia computer science*, Vol. 4, pp. 1572-1581.
- [6] Han, J., J. Pei, and M. Kamber(2011), *Data Mining: Concepts and Techniques*, Elsevier.
- [7] Harris, D. and S. Harris(2007), *Digital design and computer architecture*, Morgan Kaufmann.
- [8] Hastie, T., R. Tibshirani, and J. Friedman(2009), *The elements of statistical learning: data mining, inference, and prediction*, Springer Science & Business Media.
- [9] Jun, S. Y., Y. M. Kim, B. G. Woo, and H. Chung(2008), A Systematic Approach to Decide Maximum Berthing Ship Size Coupled with Berth Design Criteria, *Journal of the Korean Society of Marine Environment & Safety*, Vol. 14, No. 1, pp. 45-54.
- [10] Kanal, L. N.(2003), Perceptron, *Encyclopedia of Computer Science*, pp. 1383-1385.
- [11] Kim, M. K., J. H. Kim, and H. Yang(2019), Gyroscope Signal Denoising of Ship's Autopilot using Kalman Filter and Multi-Layer Perceptron, *Journal of the Korean Society of Marine Environment & Safety*, Vol. 25, No. 6, pp. 809-818.
- [12] Kohavi, R.(1995), A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, Vol. 14, No. 2, pp. 1137-1145.
- [13] Lee, S. W., J. W. Cho, and I. S. Cho(2019), Estimation of Berthing Velocity Using Probability Distribution Characteristics in Tanker Terminal. *Journal of Navigation and Port Research*, Vol. 43, No. 3, pp. 186-196.
- [14] Ministry of oceans and fisheries(2017), *Harbor and Fishery Design Criteria*.
- [15] PIANC(2020), *Berthing Velocity Analysis of Seagoing Vessels over 30,000DWT*, Working group 145 of the MARITIME NAVIGATION COMMISSION.
- [16] Rosenblatt, F.(1962). *Principles of Neurodynamics*, Spartan Books.
- [17] Roubos, A., L. Groenewegen, and D. J. Peters(2017), Berthing velocity of large seagoing vessels in the port of Rotterdam. *Marine Structures*, Vol. 51, pp. 202-219.
- [18] Shalev-Shwartz, S. and S. Ben-David(2014), *Understanding machine learning: From theory to algorithms*, Cambridge university press.
- [19] Tukey, J. W.(1977), *Exploratory Data Analysis*, Addison-Wesley Pub. Co.
- [20] Zheng, A. and A. Casari(2018), *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientist*, O'Reilly Media Inc.

Received : 2020. 03. 06.

Revised : 2020. 04. 13.

Accepted : 2020. 04. 27.