



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

工學碩士 學位論文

머신러닝 예측 알고리즘을 이용한
선박 접안속도에 영향을 미치는 요인 분석

Analysis of Factors influencing Berthing Velocity of Ship
using Machine Learning Prediction Algorithm

指導教授 趙翼順

2019 년 08 월

한국해양대학교 대학원

해양플랜트운영학과

이형탁

본 논문을 이형택의 공학석사 학위논문으로 인준함.

위원장 김태균



위원 예병덕



위원 조익순



2019년 06월 21일

한국해양대학교 대학원

목 차

List of Tables	iii
List of Figures	v
Abstract	vii
1. 서 론	
1.1 연구 배경 및 목적	1
1.2 연구 방법	3
2. 선박 접안에너지	
2.1 접안에너지	5
2.1.1 접안에너지 산정	5
2.1.2 접안에너지 영향요인 분석	6
2.2 접안속도	10
2.2.1 PIANC	11
2.2.2 국내외 접안속도 기준 및 연구사례	16
3. 빅데이터와 머신러닝	
3.1 빅데이터	21
3.2 머신러닝	25
3.3 의사결정나무	32
3.4 랜덤포레스트	35
3.5 로지스틱회귀	38
3.6 인공지능경망	40
4. 실측 데이터 분석	
4.1 데이터 수집 및 선정	43

4.1.1 데이터 수집	43
4.1.2 데이터 선정	46
4.2 기초 데이터 분석	53
4.2.1 범주형 변수	53
4.2.2 연속형 변수	62
4.3 연관성분석	66
4.3.1 상관분석	66
4.3.2 교차분석	69
4.3.3 변수채택	71
4.4 다중선행회귀	72
5. 예측모델 알고리즘	
5.1 개요	77
5.2 접안속도 예측모델	81
5.2.1 의사결정나무	81
5.2.2 랜덤포레스트	83
5.2.3 로지스틱회귀	85
5.2.4 인공신경망	86
5.3 예측모델 성능비교	90
5.3.1 혼동 행렬	90
5.3.2 ROC 곡선	94
5.4 최종모델 선택 및 영향요인 분석	96
6. 결론 및 향후과제	
6.1 결론 및 향후과제	98
참고문헌	101

List of Tables

Table 2.1	Calculation $M_s(t)$	7
Table 2.2	The factor of actual approach velocity	12
Table 2.3	The result of ship' s berthing velocity analysis in PIANC WG145	13
Table 2.4	Proposed design berthing velocity in PIANC WG145	14
Table 2.5	The factor of actual approach velocity between WG33 and WG145	15
Table 2.6	Berthing velocities depending on ship types	19
Table 3.1	Interpretation of Machine Learning	25
Table 3.2	Machine Learning algorithm	29
Table 3.4	Basic characteristic of decision tree algorithms	34
Table 4.1	Particulars of measured data	45
Table 4.2	Berth particular and operation regulations	45
Table 4.3	DAS data information	47
Table 4.4	Selection of Variables	48
Table 4.5	Terminal approaching Regulation	51
Table 4.6	Variables and descriptions	52
Table 4.7	Cross anlysis of JettyNo.-Berthing Velocity	56
Table 4.8	Cross anlysis of STATE-Berthing Velocity	58
Table 4.9	Cross anlysis of Pilot-Berthing Velocity	60
Table 4.10	Qualitative interpretation of Correlation coefficient	67
Table 4.11	Correlation Analysis of Continuous Variables	67
Table 4.12	Correlation Analysis of Categorical Variables	68
Table 4.13	Cross anlysis of JettyNo.-STATE	69
Table 4.14	Cross anlysis of JettyNo.-Pilot	70

List of Tables

Table 4.15	Cross analysis of STATE-Pilot	71
Table 4.16	Selection Variables	72
Table 4.17	Dummy Variable Transformation	74
Table 4.18	Result of Multiple Regression	75
Table 4.19	Result of Multiple Regression after stepwise selection	76
Table 5.1	Analysis environment	77
Table 5.2	Machine Learning algorithm package in R	77
Table 5.3	Hyperparameter	78
Table 5.4	Classified Train data in Decision Tree	81
Table 5.5	Classified Train data in Random Forest	83
Table 5.6	Result of Logistic Regression	85
Table 5.7	Classified Train data in Logistic Regression	86
Table 5.8	Weights of Artificial Neural Network	87
Table 5.9	Constant of Artificial Neural Network	87
Table 5.10	Discriminant Function of Artificial Neural Network	88
Table 5.11	Classified Train data in Artificial Neural Network	89
Table 5.12	2*2 Confusion Matrix	90
Table 5.13	Result of Confusion Matrix	91
Table 5.14	Evaluation Index of Machine Learning Model	92
Table 5.15	F-measure	93
Table 5.16	Adopting the final machine learning model	96

List of Figures

Fig. 1.1	Flow Chart of Study	4
Fig. 2.1	Geometric parameters of ship' s berthing	8
Fig. 2.2	Brolsma' s chart	16
Fig. 2.3	British Standard chart and berthing velocities	17
Fig. 2.4	Spanish ROM berthing velocities	18
Fig. 2.5	Japanese Technical Standard chart and berthing velocity	19
Fig. 2.6	Berthing velocity at Crude Oil Tanker Berth (Keiyo Sea Berth in 1971~73)	20
Fig. 3.1	From 3Vs, 4Vs, 5Vs and 6Vs Big Data characteristics	22
Fig. 3.2	3 ² Vs Venn diagrams in hierarchical model	23
Fig. 3.3	Correlation of 3 ² Vs to machine learning Venn diagrams	24
Fig. 3.4	Relationship diagram of Big Data, Machine Learning and Deep Learning	26
Fig. 3.5	Data analysis process	31
Fig. 3.6	Example of Decision Tree	32
Fig. 3.7	Bagging Algorithm	36
Fig. 3.8	Flow of Random Forest algorithm	37
Fig. 3.9	Example of Logistic curve	38
Fig. 3.10	Structure of ANN	41
Fig. 4.1	Arrangement of Dock mounted docking aid system	43
Fig. 4.2	SmartDock system	44
Fig. 4.3	Weather(wind) data	46
Fig. 4.4	Selection of Variables in PIANC WG145	50
Fig. 4.5	Berthing Velocity Frequency Analysis	53
Fig. 4.6	Jetty No. Frequency Analysis	54

List of Figures

Fig. 4.7 STATE Frequency Analysis	55
Fig. 4.8 Pilot Frequency Analysis	56
Fig. 4.9 Jetty No.(Under10) Frequency Analysis	57
Fig. 4.10 Jetty No.(Over10) Frequency Analysis	58
Fig. 4.11 STATE(Under10) Frequency Analysis.....	59
Fig. 4.12 STATE(Over10) Frequency Analysis.....	60
Fig. 4.13 Pilot(Under10) Frequency Analysis.....	61
Fig. 4.14 Pilot(Over10) Frequency Analysis.....	62
Fig. 4.15 DWT-Velocity Scatter Graph	63
Fig. 4.16 MaxAngle-Velocity Scatter Graph	64
Fig. 4.17 Wind-Velocity Scatter Graph	65
Fig. 5.1 Hold-Out method.....	79
Fig. 5.2 Result of Decision Tree	82
Fig. 5.3 Result of Random forest	83
Fig. 5.4 Importance of Dependent variables in Random forest	84
Fig. 5.5 Result of Artificial Neural Network	86
Fig. 5.6 Importance of Independent variable in Artificial Neural Network	89
Fig. 5.7 ROC curve.....	94
Fig. 5.8 Compared with ROC curve.....	95

Analysis of Factors influencing Berthing Velocity of Ship using Machine Learning Prediction Algorithm

Lee, Hyeong Tak

Department of Offshore Plant Management
Graduate School of Korea Maritime and Ocean University

Abstract

When a ship is in contact with the dock facilities of the port, the ship operator and the port manager shall consider the ship's berthing energy. Ships must meet within the allowable berthing energy of port facilities to protect the hull and prevent damage to ports facilities. The factor that has the greatest effect on the berthing energy is the berthing velocity, so it is very important for ships to meet and maintain proper berthing velocity.

The ship's berthing velocity is determined by various factors including weather, the location of the port and dock, the type of ships and human factors. However, based on the domestic and overseas berthing velocity criteria that had been studied and analyzed, the proper berthing velocity was proposed based on the collected vessel berthing data, considering only the size of the vessel. In addition, the analysis results using only statistical techniques were presented based on data.

In this study, I intend to use machine learning techniques, which are one of big data analyses, not just statistical techniques, based on vessel berthing velocity data, to suggest proper berthing velocity. Big data analysis is a method of comprehensively analyzing various variables to predict the future and make decisions, and it can be analyzed and predicted using all the various factors of ship's berthing velocity. Therefore, the vessel operator can prevent accidents by predicting the appropriate berthing velocity, and from the point of view of the port manager, it can be a safety management standard.

In this study, a prediction model was built by applying machine learning techniques based on 206 actual data collected for 17 months from March 2017 to July 2018 through a ship berthing assist device at the local tanker dock. Independent variables and dependent variables were selected by data needed for analysis from the measured data, and significant variables were adopted through correlation analysis and cross analysis.

Decision Tree, Random Forest, Logistic Regression, and Artificial Neural Network were used as machine learning algorithms. The Hold-Out method was adopted as a way to verify the accuracy of the prediction model. The Train data set and Test data set were separated by 7:3.

Prediction model built through the Train data set was used to compare performance with the indicators according to the Confusion matrix and ROC curve by using the test data set. As a result, random forest performed best and decision trees and logistic regression performed relatively low. According to Random Forest, which showed good performance, factors affecting the berthing velocity were shown in order of the loading condition of the cargo, tonnage of the ship, location of the jetty and approach angle of the vessel.

In this study, the machine learning technique, which is a big data analysis method, was applied based on actual data of the vessel's berthing velocity. As a result of the analysis, it was confirmed that the berthing velocity prediction is possible through the analysis of big data, and that the

factors affecting the berthing velocity could be identified to help prevent accidents for ship operators, port and dock managers. Through this study, it was shown that it is possible to utilize prediction model and decision-making model in the field of ship and marine transportation by using machine learning and artificial intelligence based on big data. However, further research is needed through data collection at various types of vessel and docks, and all factors affecting the berthing velocity have not been utilized in the analysis. It also needs to collect data over a long period of time to build more accurate prediction models.

KEY WORDS: Berthing Velocity of Ship, Measured Data, Big Data, Machine Learning, Prediction Algorithm.

머신러닝 예측 알고리즘을 이용한 선박 접안속도에 영향을 미치는 요인 분석

이 형 탁

한국해양대학교 대학원
해양플랜트운영학과

초 록

선박이 항만·부두시설에 접안할 때 선박운항자 및 항만관리자는 선박의 접안에너지를 고려해야한다. 선박은 항만시설의 허용 접안에너지 이내로 접안해야 선체를 보호하고 항만시설의 파손을 예방할 수 있다. 이러한 접안에너지에 가장 큰 영향을 미치는 요소는 접안속도로서 선박은 적정 접안속도를 유지하며 접안하는 것이 매우 중요하다.

선박 접안속도는 날씨, 항만의 위치, 선박의 종류, 인적요인 등 다양한 요인들의 영향을 받아 결정된다. 하지만 그동안 연구 및 분석되었던 국내외 접안속도 기준은 수집된 선박 접안데이터를 바탕으로 선박의 크기만을 고려하여 적정 접안속도를 제안하였다. 또한 데이터를 기반으로 통계적 기법만을 활용한 분석결과를 제시하였다.

본 연구에서는 선박 접안속도 데이터를 바탕으로 통계적 기법만이 아닌 빅데이터 분석 중 하나인 머신러닝 기법을 활용하고자 한다. 빅데이터 분석은 다양한 변수를 종합적으로 분석하여 미래를 예측하고 의사결정을 하는 방법으로서 선박 접안속도의 다양한 영향요소를 모두 활용하여 분석하고 예측을

할 수 있다. 따라서 선박운항자는 접안 시 적정 접안속도를 예측하여 사고 예방을 할 수 있으며, 항만관리자 입장에서는 안전 관리에 참고할 수 있다.

따라서 본 연구에서는 국내의 한 탱커부두에서 선박 접안 보조 장치를 통해 약 17개월(2017.03. ~ 2018.07.) 간 수집한 206개의 실측데이터를 바탕으로 머신러닝 기법을 적용하여 예측모델을 구축하고자 한다. 수집한 실측데이터에서 분석에 필요한 데이터를 선별하여 독립변수와 종속변수를 선정하고 상관관계 분석 및 교차분석을 통해 유의미한 변수를 채택하였다.

머신러닝 알고리즘으로는 의사결정나무(Decision Tree), 랜덤포레스트(Random Forest), 로지스틱 회귀분석(Logistic Regression), 인공신경망(Artificial Neural Network)을 활용하였다. 예측모델의 예측정확도 검증을 위한 방식으로는 Hold-Out 방식을 채택하였으며 Train data set과 Test data set을 7:3으로 분리하였다.

Train data set을 통해 구축된 예측모델은 Test data set을 이용하여 혼동행렬에 따른 지표와 ROC 곡선으로 성능을 비교하였다. 그 결과, 랜덤포레스트가 가장 좋은 성능을 보였으며 의사결정나무, 로지스틱 회귀분석이 비교적 낮은 성능을 보였다. 좋은 성능을 보인 랜덤포레스트에 따르면 접안속도에 영향을 미치는 요인은 화물의 적재상태, 선박의 질량, 접안 부두 위치, 접안 각도 순으로 나타났다.

본 연구에서는 선박 접안속도 실측데이터를 바탕으로 빅데이터 분석방법인 머신러닝 기법을 적용하였다. 분석 결과, 빅데이터 분석방법을 통하여 접안속도 예측이 가능하다는 것을 확인하였으며, 접안속도에 영향을 미치는 요인을 식별하여 선박운항자와 항만관리자의 사고 예방에도 도움을 줄 수 있음을 확인하였다. 이를 통해 선박 및 해운분야에서도 데이터 축적을 통하여 빅데이터에 기반한 머신러닝, 인공지능 등의 기법으로 예측모델, 의사결정모델 등의 활용 가능성을 제시하였다. 하지만 다양한 선종 및 부두의 데이터 수집을 통한 추가 연구가 필요하며, 접안속도에 영향을 미치는 모든 요인을 분석에 활용하지 못한 한계점을 가진다. 또한 오랜 기간동안 데이터를 수집하여 더욱 정확한 예측모델을 구축할 필요가 있다.

KEY WORDS: 선박 접안속도, 실측데이터, 빅데이터, 머신러닝, 예측 알고리즘.

제 1 장 서 론

1.1 연구 배경 및 목적

기술의 발달로 수많은 시스템, 센서, 장치에 의해 생성되는 데이터의 양이 급격히 증가하고 있으며 2020년에는 전 세계 디지털 데이터의 양이 40조 기가바이트에 이를 것으로 추정하고 있다(Gantz et al., 2012). 이는 제4차 산업혁명 시대를 맞이하게 되는 계기가 되었으며, 이를 이끄는 분야는 인공지능, 사물인터넷, 자율주행 자동차, 로봇, 공유경제, 3D 프린터 등 빅데이터(Big data)에 기반한 정보통신기술과 관련이 있다. 이제 빅데이터는 일상생활을 포함한 거의 모든 영역에서 유용하게 활용되어지고 있으며 앞으로도 더 넓은 영역에서 더욱 세세한 문제에 이르기까지 그 적용이 확장될 것임이 자명하다(김진호, 2016).

빅데이터의 정의는 초기에 기존의 소프트웨어가 처리할 수 없을 정도로 크고 복잡한 데이터 세트를 설명하기 위해 사용되었다(Snijders, et al., 2014). 하지만 오늘날의 빅데이터는 대량의 데이터를 다루는 문제를 넘어서는 개념이며, 데이터 자체뿐만 아니라 그러한 데이터가 제공하는 기회를 활용하기 위한 기술 동향의 진보를 의미하며 이는 미래를 예측하고 의사결정을 내리는 새로운 접근법이다.

구글트렌드에 따르면 ‘빅데이터’의 관심도는 구글트렌드 서비스가 시작된 2004년에는 0에 가까웠으나 2016년부터 현재에는 관심도가 100에 육박하고 있다. 또한 세계경제포럼에서 ‘2012년의 떠오르는 10대 기술(Top 10 Emerging Technologies of 2012)’ 중 하나로 빅데이터를 지목한 이후 전 산업권에서 빅데이터의 관심도는 꾸준히 증가하였다. 시장조사기관인 Wikibon에 따르면 소프트웨어, 하드웨어, 서비스를 모두 포함한 세계 빅데이터 시장은 2026년에 이르러서는 총 922억 달러의 규모로 성장할 것으로 전망했다(박선우, 2018).

이러한 동향과는 다르게 선박 및 해운분야에서는 빅데이터를 활용한 연구가 활발하지 않았다. 이는 선박 간 통신 한계와 해운 분야의 네트워크 규모 때문

에 전통적으로 정보 집약적이지 않았고, 이에 따라 데이터 축적 부족으로 인하여 데이터를 활용할 수 없었기 때문이다(Maris Mirovic et al., 2018).

하지만 V-sat 등 기기의 발달로 인하여 선박 간 통신문제가 해결되면서 선박 및 해운분야에서는 빅데이터 활용의 중요성을 인지하고 지난 몇 년간 데이터를 축적하기 시작했다. 선박의 GPS, AIS, ECDIS 등 여러 항해기기에서 제공되는 운항 데이터, 항만 및 부두에서 수집되는 접안 데이터, 화물의 선적·하역 데이터 등은 선박 및 해운분야의 사고 예방 뿐 아니라 미래의 예측 및 의사결정을 하는데 기초자료로서의 활용가치가 높다.

최근 국제적으로 DAS(Dock mounted docking aid systems)와 같은 선박 접안 보조 장치들의 중요성이 부각되면서, 많은 항만 부두 시설에 접안속도 측정 장비들이 설치되어 접안 실측데이터를 축적하고 있다. 이는 선박이 항만 부두 시설 접안할 때 발생하는 접안에너지 구성요소 중, 접안속도가 가장 큰 영향을 주는 요소이기 때문이다. 선박의 접안속도가 과도한 상태로 접안하게 되면 부두의 계류시설이 파괴되거나 해당 선박의 선체 손상과 같은 사고로 이어지게 되는데 이러한 사고 발생 시, 부두 및 선박의 수리비는 물론 부두의 잠정적 폐쇄로 인한 경제적인 피해가 발생하게 된다(이상원, 2019).

국내에서도 법률을 통해 접안 보조 장치의 설치가 권장되어 대부분의 위험물 운반선의 주요 부두에 설치가 되고 있으나, 다른 선종의 부두에는 접안 보조 장치들의 설치가 되지 않고 있다. 접안속도의 중요성으로 인하여 PIANC 등 국외 연구단체에서는 적정 접안속도 기준을 제시하고 있지만 국내는 접안속도에 대한 관련 연구는 부족한 상황이며, 대부분 외국의 기준 및 연구사례를 그대로 번역하여 사용하고 있다. 따라서 국내에서도 접안속도에 대한 데이터를 취합, 분석하여 부두 계류시설 및 선박 사고 예방에 대한 연구가 필요하다(이상원, 2019).

이러한 국제적 추세에 따라 본 연구에서는 접안속도의 중요성을 인식하고 부두 계류시설 및 선박 사고 예방을 위해 접안에 필요한 요인을 상호 분석하여 어떠한 요인이 접안속도에 가장 영향을 미치는지 파악하고자 한다. 분석방법으

로는 DAS와 같은 선박 접안 보조 장치에서 제공되는 선박 접안 실측데이터를 바탕으로 빅데이터 분석을 이용하였다. 빅데이터 분석 방법 중 머신러닝(Machine Learning) 기법을 사용하여 실측데이터를 분석하고, 적정 접안속도를 기준으로 초과유무 예측알고리즘을 구축하고자 한다. 이는 선박 접안 시 적정 접안속도를 예측하고 접안속도에 영향을 미치는 요인을 분석함으로써 접안 사고 예방에 도움이 될 것이라 사료된다. 또한 선박 및 해운분야의 빅데이터 활용법을 제시하여 활용할 수 있는 기초 자료를 제공하고자 한다.

1.2 연구 방법

본 논문은 6개의 장으로 구성되어있다.

제 2장은 접안에너지를 산정하는 방법 및 접안에너지 영향요인을 분석하였다. 접안에너지의 영향요인 중 접안속도의 중요성을 기술하였으며, 국내외 접안속도와 관련된 연구와 기준에 대한 검토를 통해 비교 분석하였다.

제 3장에서는 본 연구의 분석기법인 빅데이터와 머신러닝에 대한 내용을 조사하였으며 선박 및 해운분야에서의 빅데이터 활용 연구사례를 검토하였다. 또한 머신러닝 기법 중 지도학습 알고리즘인 의사결정나무(Decision Tree), 랜덤 포레스트(Random Forest), 로지스틱회귀(Logistic Regression), 인공신경망(Artificial Neural Network)을 소개한다.

제 4장에서는 접안속도 실측데이터에 대한 수집 및 선정과정을 설명하고, 기초적인 데이터 분석을 실시하였다. 또한 데이터 간 연관성 분석을 통해 변수를 선택하고 다중선형회귀 분석을 실시하였다.

제 5장에서는 접안속도 예측모델 알고리즘을 구축하였다. 그리고 혼동행렬과 ROC 곡선을 통한 성능 평가 지표를 비교·분석하여 최적의 알고리즘을 선정한다. 그리고 최적의 알고리즘에 따른 접안속도에 가장 영향을 미치는 요인을 파악한다.

제 6장에서는 본 논문의 접안속도 예측모델 알고리즘으로부터 얻어진 결과에 대해 요약하고 향후 연구과제에 대하여 기술하였다.

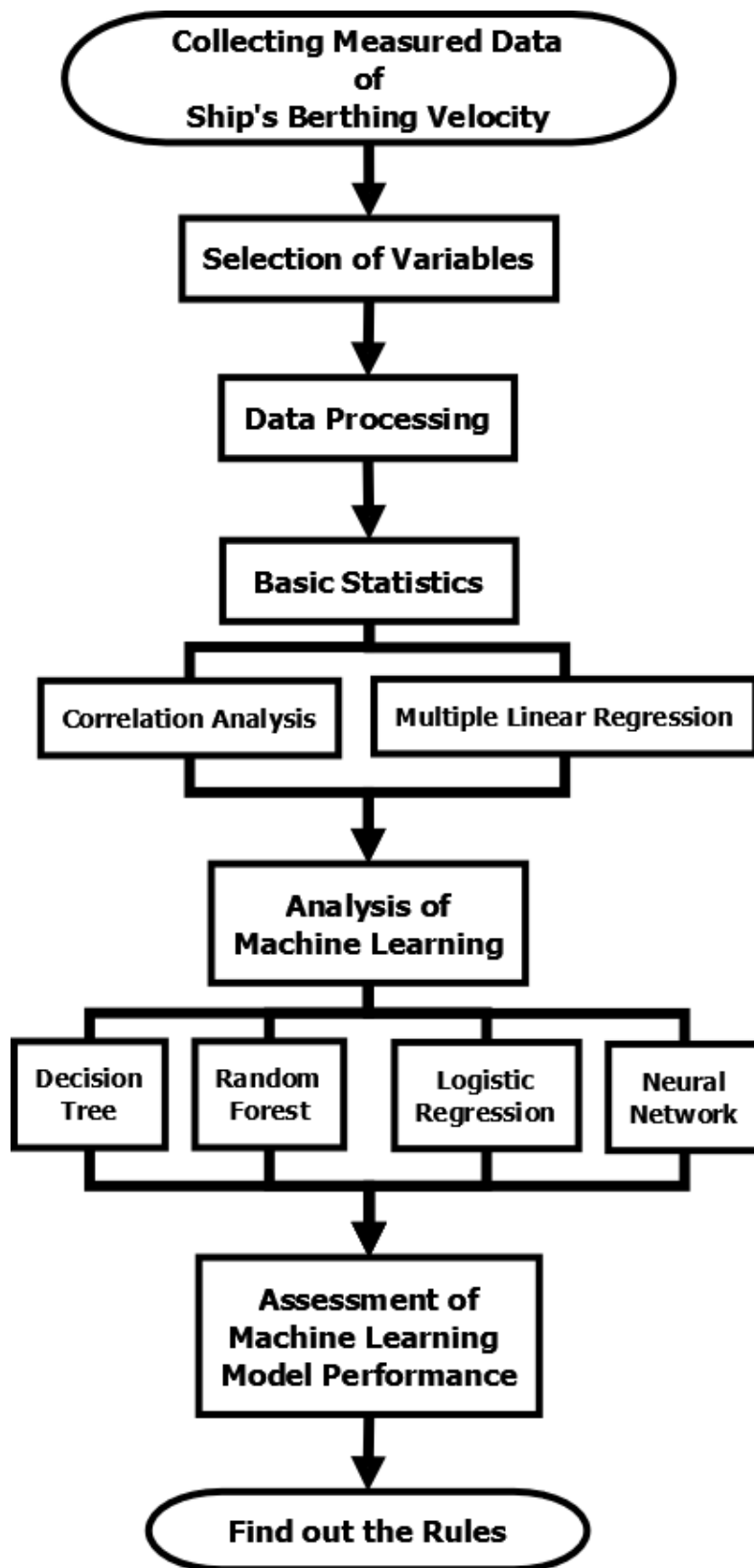


Fig. 1.1 Flow Chart of Study

제 2 장 선박 접안에너지

2.1 접안에너지

선박의 접안에너지를 구하는 방법에는 통계적 방법, 수리모형실험에 의한 방법, 유체역학적 모델, 운동역학적방법 등이 있으나 항만 및 부두 설계에 필요한 데이터가 부족하거나 계산을 위한 여러 가지 정수가 충분히 밝혀지지 않았기 때문에 운동역학적인 방법을 일반적으로 사용하게 된다(항만 및 어항 설계기준, 2017).

2.1.1 접안에너지 산정

선박 접안에너지를 산정하는데 사용되는 운동역학적방법은 일반적이고 보편적으로 사용되는 방법이다. 국내의 항만 및 어항 설계기준에는 운동역학적방법으로 선박의 접안에너지 계산법을 제시하고 있다. 선박 운동에너지는 다음 식 (2.1)과 같다.

$$E_s = \frac{1}{2} M_s V^2 \quad (2.1)$$

E_s : 선박의 운동에너지(kN · m)

M_s : 선박의 질량 (t)

V : 선박의 접안속도 (m/s)

하지만, 실제 선박 접안 시 접안에너지는 다음과 같은 요소를 고려하여 접안 에너지를 산정한다.

- 선박의 회전
- 접안 구조물의 탄성적 변형

- 방충시스템의 압출
- 선박과 접안 구조물 사이의 배수량

위의 고려사항들을 포함한 선박 접안에너지 산정식은 식 (2.2)와 같다.

$$E_f = \frac{1}{2} M_s V^2 C_e C_m C_s C_c \quad (2.2)$$

E_f : 선박의 접안에너지(kN · m)

M_s : 선박의 질량 (t)

V : 선박의 접안속도 (m/s)

C_e : 편심계수 (Eccentricity coefficient)

C_m : 가상질량계수 (Mass coefficient)

C_s : 유연성계수 (Softness factor, 표준은 1.0)

C_c : 선석의 형상계수 (Berth configuration factor, 표준은 1.0)

2.1.2 접안에너지 영향요인 분석

선박 접안에너지 계산식(2.2)에 따르면 운동에너지 공식에 따른 선박의 질량과 접안속도, 그리고 4가지의 계수에 의해 계산되어진다.

선박의 질량은 국내 항만 및 어항 설계기준에 따르면 재화중량톤수(DWT), 또는 총톤수(GT)와 만재시의 배수량(DT)의 관계식을 사용한다. 다음 Table 2.1은 그 관계식을 나타낸다.

Table 2.1 Calculation $M_s(t)$

Ship	Calculation
Cargo ship	DT = 1.174DWT
Container	DT = 1.385DWT
Tanker	DT = 1.235DWT
Ro-Ro	DT = 1.022GT
Car carrier	DT = 0.751GT
LPG	DT = 1.400GT
LNG	DT = 1.118GT
Passenger	DT = 0.573GT
Ferry (long distance)	DT = 1.240GT
Ferry (short distance)	DT = 1.279GT

Source: 항만 및 어항 설계기준, 2017

접안 시 선박은 선석과 평행하지 않게 되는데, 방충재(Fender)와 접촉한 후 회전(yawing)을 하며 그 결과 소모되게 되는 에너지에 대한 계수가 C_e (편심계수, Eccentricity coefficient)이다. 편심계수를 계산하는 방법은 Fig 2.1과 식(2.3) ~ (2.6)과 같다(이상원, 2019). 편심계수는 선박의 접안각도에 영향을 받으므로 적정 접안각도 제시 및 접안각도에 대한 정확한 분석이 필요하다.

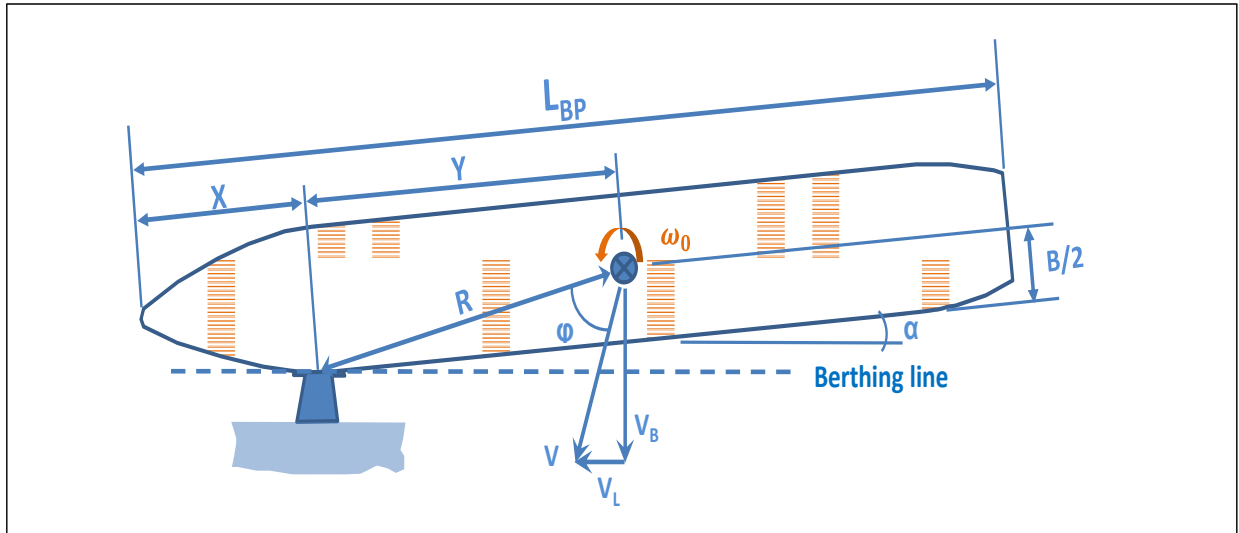


Fig. 2.1 Geometric parameters of ship's berthing

$$R = \sqrt{\left(\frac{L}{2} - x\right)^2 + \frac{B^2}{4}} \quad (2.3)$$

$$K = (0.19C_b + 0.11)L_{pp} \quad (2.4)$$

$$\Phi = 90 - \alpha - \alpha \sin\left(\frac{B}{2R}\right) \quad (2.5)$$

$$C_e = \frac{K^2 + R^2 \cos(\Phi)}{K^2 + R^2} \quad (2.6)$$

R : 선박의 무게중심과 선박-구조물 사이의 충돌 지점 사이의 반경

(The radius between the centre of mass of the ship and the point of collision between the ship and the structure, m)

K : 선박의 선회반경 (Radius of gyration of the ship, m)

α : 접안각도 (berthing angle, deg)

C_m (가상질량계수, Mass coefficient)는 선박이 접안 시 선박 주변의 바닷물이 같이 움직이게 되는데 이 바닷물의 양을 선박의 질량에 포함시키는 계수이다. 이 가상질량계수는 식 (2.7)과 같이 나타낸다.

$$C_m = \frac{M_s + M_w}{M_s} \quad (2.7)$$

M_s : 선박의 질량 (선박의 배수량/중력가속도)

M_w : 선박주변의 바닷물 질량 (부가질량, t)

일반적으로 쓰이는 가상질량계수의 계산방식은 2가지가 있다. Ueda (1981)의 방식과 Costa (1964)의 방식이 있다. Ueda (1981)의 방식으로 가상질량계수를 구하는 식은 다음의 (2.8), (2.9)의 식과 같다.

$$C_m = 1 + \frac{\pi}{2C_b} \times \frac{d}{B} \quad (2.8)$$

$$C_b = \frac{\nabla}{L_{pp}Bd} \quad (2.9)$$

C_b : 방형비척계수 (Block coefficient)

∇ : 선박의 배수부피 (Displacement, m³)

L_{pp} : 수선간 길이 (m)

B : 형폭 (m)

d : 만재흘수 (m)

이와 같은 Ueda의 방식은 국내의 항만 및 어항기준(2014)에서 채택하는 방식이다. 다른 방식인 Costa (1964)의 방식은 식 (2.10)과 같다.

$$C_m = 1 + \frac{2d}{B} \quad (2.10)$$

선박이 접안할 시, 계류시설(방충재)에 접촉하게 되며 이로 인하여 선체외판의 변형이 일어나 에너지가 흡수된다. 이를 C_s (선박의 유연성계수, Softness factor)라고 하며, 방충재가 유연한 경우 일반적으로 C_s 의 값은 1.0으로 취한다.

C_c (선석의 형상계수, Berth configuration factor)는 선박이 접안 시 선박과 계류시설 사이에 있는 바닷물이 압축되어 완충효과를 내게 된다. 이 때 방충재에 흡수되는 에너지를 감소시키는 효과가 나타나게 되는데 보통 C_c 의 값은 1.0으로 취한다. 이 현상은 접안각도, 선박외판의 형상, Under keel clearance 및 접안속도 등에 관계되는 것으로 보이나 아직 해명되지 않은 부분이 많다(항만 및 어항 설계기준, 2017).

접안속도(V)는 선박 접안에너지의 운동역학적 산정 방식에 따르면 제곱값을 가지므로 접안에너지 산정에 영향을 미치는 가장 큰 요인이라고 할 수 있다. 따라서 적정 접안속도를 유지하며 접안하는 것이 해당 부두 설계 선박 접안에너지에 충족할 수 있다. 따라서 과도한 접안속도를 가지고 접안을 하게 되면 선박 및 항만, 부두의 사고로 이어질 수 있다는 것을 의미한다.

따라서 본 연구에서는 선박 접안에너지의 운동역학적 산정 방식에 의해 접안속도의 중요성을 인지하고 DAS(Dock mounted Aid System)에서 측정된 선박 접안데이터를 분석하고 머신러닝 기법을 이용하여 예측 알고리즘을 설계하고자 한다.

2.2 접안속도

선박의 접안에너지에 가장 영향을 미치는 요소로는 접안속도이다(PIANC, 2012). 따라서 선박 접안 시 적절한 접안속도를 유지하며 접안하는 것이 중요하다. 접안속도는 자연외력에 노출되어있는지 유무, 접안하기에 어려운지 혹은 비

교적 쉬운지 등과 같은 해당 부두의 특성에 따라 적정 접안속도가 정해진다(조익순 등, 2018).

선박이 접안함에 있어 접안속도의 중요성으로 인하여 국내외에서는 접안속도 기준을 제시하고 있다. 국내에서는 주로 일본의 기준을 그대로 번역하여 사용하고 있는 경우가 대부분이다. 하지만 일본, 유럽 등에서는 접안속도에 관한 연구가 활발히 진행되고 있으며 이에 대한 기준을 제시하고 있다(이상원, 2019).

2.2.1 PIANC

PIANC(The World Association for Waterborne Transport Infrastructure)는 1885년에 설립되었으며 항만, 수송, 선박 등에 관련된 기술적, 경제적, 환경적 문제에 대해 공공 및 민간에서 최고의 국제적 전문가들을 모아 지침과 기술적 조언을 제안하는 단체이다. PIANC에서 발행한 Technical Report 중 WG33 (Working Group 33, Guidelines for the design of fender system, 2002), WG145 (Working Group 145, Draft Report of Berthing velocities in Sheltered Environment, 2017)에서 접안속도의 기준을 제시하고 있다.

항만 부두의 방충재(Fender) 설계에 대한 Technical Report인 WG33에서는 설계 접안속도에 대한 기준을 제시한다. PIANC WG33에서는 접안속도가 접안에너지를 계산하는데 있어 가장 영향을 크게 미치는 요소이며, 접안속도를 결정하는데 영향을 미치는 요인을 Table 2.2와 같이 나타낸다.

PIANC WG33(2002)에서는 방충재의 설계를 위한 선박의 접안속도를 검토하였다. 그러나 최근 컨테이너선과 벌크선 등 선박의 급속한 대형화와 확대에 인하여 방충재의 설계에 대하여 기준을 다시 제시할 필요가 있으며, 이에 가장 큰 영향을 미치는 접안속도에 대한 기준을 제안할 필요성이 대두되었다. 따라서 PIANC는 2010년 WG 145를 설립하여 여러 국가의 다양한 부두에서 접안속도를 측정하여 연구, 재검토가 이루어지고 있으며 2017년 Draft Report를 발표하였다.

Table 2.2 The factor of actual approach velocity

Factor	Remark
Prevailing physical boundary conditions	the influence of waves, wind and current should be considered
Ease of navigation	is the approach to the berth easy or difficult?
Method of berthing	are berthing aids used, is berthing always parallel, when is the forward motion of the vessel stopped, etc
Type of vessel	is the vessel equipped with powerful engines, quick reacting engines, bow thrusters, etc
Use of tugs	are tug boats used, how many and of sufficient capacity?
Frequency of berthing	at berths with a high berthing frequency, generally higher berthing velocities are experienced
Size of Vessel	the approach velocity of larger vessels is generally less than the approach velocity of smaller vessels; range of vessels expected at the berth must be considered
Berth appearance	ship masters will berth more careful when approaching a desolate berth instead of a new, modern berth
Type of cargo	a vessel with hazardous cargo will generally berth under better controlled circumstances, the use of berthing aids for example
Windage area of the vessel	a vessel with a windage area is considerably more susceptible to wind
Human factor	a most important factor, this may concern the level of experience, etc

Source : PINAC Working Group 33, 2002

PIANC WG145(2017)에서는 접안속도 실측치를 분석하여 부두의 특성, 운항자의 경험, 자연 환경적 요인에 따라 Group A와 Group B로 구분하여 선박규모와 초과확률에 대한 접안속도 예측치를 Table 2.3과 같이 분석하였으며, Table 2.4와 같이 부두 설계기준속도를 제시하였다.

Table 2.3 The result of ship' s berthing velocity analysis in PIANC WG145

m/s	Group A (over 30,000DWT ship)		Group B (over 30,000DWT ship)	
	Non-Hazardous Cargo	Hazardous Cargo	Non-Hazardous Cargo	Hazardous Cargo
Average	0.040	0.050	0.071	0.076
Standard Deviation	0.023	0.022	0.042	0.038
Coefficient of Variance	0.58	0.44	0.58	0.49
95% Confidence Value	0.085 (0.062~0.099)	0.092 (0.076~0.101)	0.152 (0.150~0.179)	0.148 (0.142~0.155)
99% Confidence Value	0.122 (0.086~0.146)	0.122 (0.093~0.139)	0.219 (0.210~0.243)	0.202 (0.196~0.210)
Max observed Value	0.140	0.170	0.246	0.320
Max DWT	364,767	320,780	194,435	164,533
Data Number	422	377	854	164
Group A : well-organized berthing manourver Group B : normal berthing manourver				

Source : PIANC WG145, 2017

Table 2.4 Proposed design berthing velocity in PIANC WG145

Vessel Type Group		Berthing Velocity	Confidence level
Non-Hazardous Cargo	Group A	0.09m/s	95% Confidence Value
	Group B	0.15m/s	
Hazardous Cargo	Group A	0.13m/s	99% Confidence Value
	Group B	0.20m/s	
Group A : well-organized berthing manourver Group B : normal berthing manourver			

Source : PIANC WG145, 2017

Table 2.4와 같이 PIANC WG145(2017)에서 제안한 설계 접안속도는 일반 화물선의 경우 95%의 신뢰값에 대한 접안속도를 적용하며, 탱커선의 경우 취급화물의 위험성을 고려하여 99%의 신뢰값에 대한 접안속도를 적용할 수 있다고 기술하였다. 그리고 10,000DWT 이하의 선박이거나, 예인선이 없이 접안하는 선박에는 적용되지 않는다.

또한 PIANC WG145(2017)에서는 WG33(2012)에 접안속도에 영향을 미치는 요소에 접안 보조 장치인 ‘Berthing Aids’ 라는 항목을 추가하였다. WG33(2012)와 WG145(2017)의 접안속도에 영향을 미치는 요소를 비교 검토한 표는 Table 2.5와 같다(PIANC, 2017; 이상원, 2019).

Table 2.5 The factor of actual approach velocity between WG33 and WG145

Category	Factor (WG33)	Factor (WG145)	Remark
Environment	Prevailing physical boundary conditions	Current	the influence of waves, wind and current should be considered
		Effect waves	
		Effect wind	
	Berth appearance	Closed/Open quay/jetty	ship masters will berth more careful when approaching a desolate berth instead of a new, modern berth
	Ease of navigation	-	approach to the berth easy or difficult?
-	<u>Berthing aids</u>	Available(PPU, sonar, fixed-laser) and used as a tool for berthing assistance	
Vessel	Type of vessel	Vessel type	vessel equipped with powerful engines, quick reacting engines, bow thrusters, etc.
	Size of vessel		approach velocity of larger vessels is generally less than the approach velocity of smaller vessels; range of vessels expected at the berth must be considered.
	Type of cargo		vessel with hazardous cargo will generally berth under better controlled circumstances, the use of berth aids for example
	Windage area of vessel		vessel with a large windage area is considerable more susceptible to wind
Human	Method of berthing	Berthing manoeuvre	berthing always parallel, when is forward motion of the vessel stopped, etc.
	Use of tugs	Equipment / tugs	tug boats used, how many and of sufficient capacity?
	Human factor	Human factor	most important factor, this may concern the level of experience, etc.

Source : 이상원, 2019

2.2.2 국내외 접안속도 기준 및 연구 사례

Brolsma's chart는 WG33(2002), WG145(2017) 뿐 아니라 전 세계적으로 접안속도 기준 설정에 참조되는 접안속도와 선박의 DWT 사이의 관계를 나타낸 그래프이다. 특히 British standard(BS 6349-4, 2014)에서는 Brolsma's chart를 채택하여 적정 접안속도를 제안하였고, PIANC에서는 Brolsma's chart의 평균 접안속도 값은 50% 신뢰 수준과 동일하다고 간주하였다. Fig 2.2은 Brolsma's chart로서 접안조건과 부두의 위치에 따른 5개의 곡선을 포함한다.

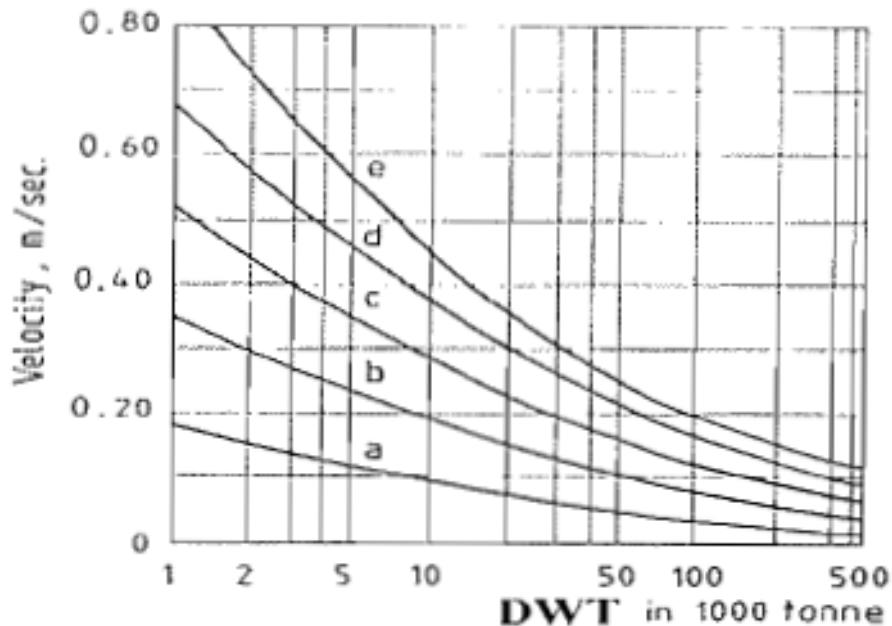
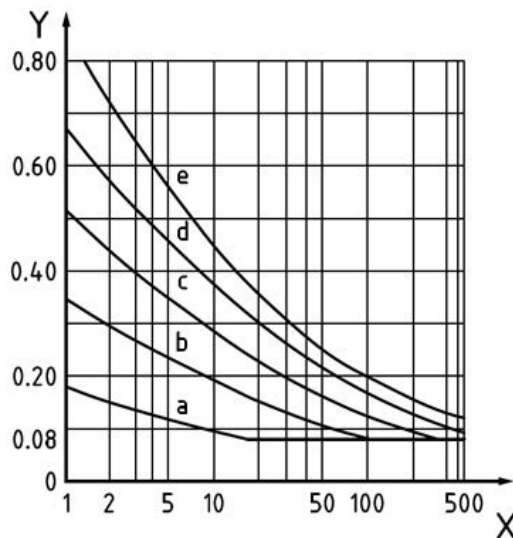


Fig. 2.2 Brolsma's chart

- a : good berthing conditions, sheltered
- b : difficult berthing conditions, sheltered
- c : easy berthing conditions, exposed
- d : good berthing conditions, exposed
- e : navigation difficult conditions, exposed

Source : Brolsma, 1977

영국의 British Standard는 Brolsma's chart를 기초로 하여 선박의 배수량과 접안조건과의 관계를 적정 접안속도 기준에 반영하였다. 특히, 적정 접안속도로 실제 접안 상태와 상관없이 0.08m/s를 권고하고 있다. Fig. 2.3는 British Standard의 접안속도 기준을 나타낸다.



Key

X Water displacement, in 1 000 t

Y Characteristic velocity, in m/s, perpendicular to the berth

Fig. 2.3 British Standard chart and berthing velocities

Source : British Standard, 2014

Beckett Rankine (2010)은 과거의 자료인 Brolsma's chart가 통계적으로 맞지 않으며 현재의 선박에는 적용하기 어려워 개정할 필요가 있다고 주장하였다. 이에 PIANC 및 British Standard에서도 선박 접안속도의 새로운 기준에 대한 연구를 진행하고 있다.

Spanish ROM의 경우 마찬가지로 선박의 배수량과 접안조건과의 관계를 바탕으로 기상조건(바람, 파도, 조류)에 따라 favourable conditions, intermediate conditions, very unfavorable conditions의 3개의 곡선으로 접안속도의 기준을 나타내었다. 이에 따른 Chart는 Fig. 2.4와 같다.

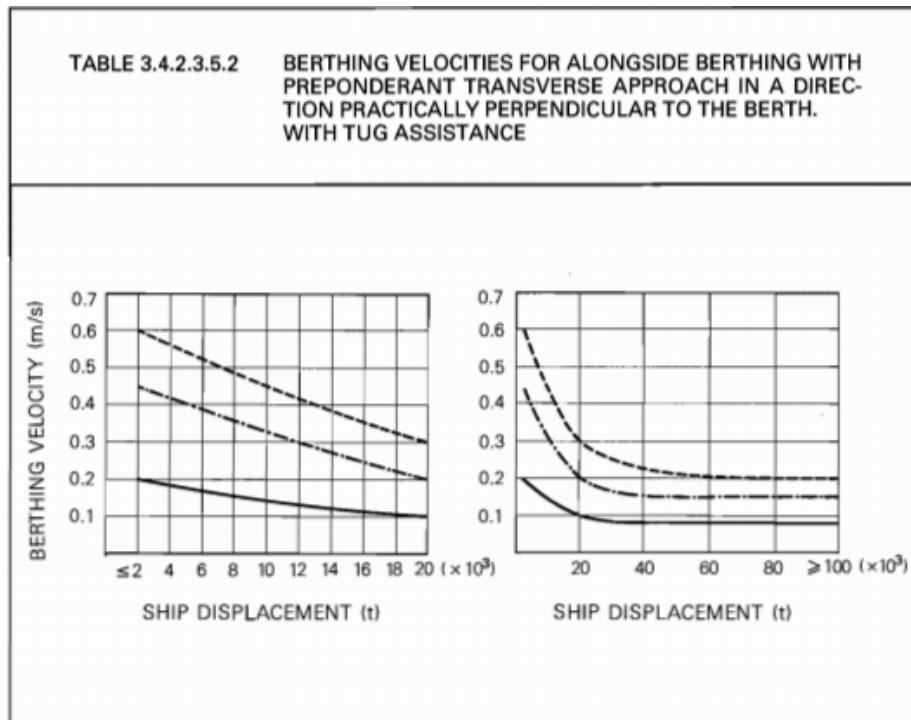


Fig. 2.4 Spanish ROM berthing velocities

Source : Spanish ROM 3.1-99

Roubos et al. (2016)은 네덜란드 로테르담 항의 실측 접안데이터를 바탕으로 접안속도에 영향을 미치는 요소를 식별하였다. 선종별, 규모별로 구분하여 접안속도를 분석하였고, UKC, 자연 환경, 예인선의 수 등이 접안속도에 미치는 영향을 연구하였다. 그 결과 접안속도는 접안 관련 정책, 도선사의 숙련도, 접안 보조 장치 등에 큰 영향을 받는다고 주장하였다.

Japanese Technical Standard는 실증적 데이터를 바탕으로 적정 접안속도를 제안하였다. Japanese Technical Standard에 따른 접안속도 제안 Chart(Fig. 2.6)는 접안조건과 부두 위치에 따른 6개의 곡선으로 나타내었다. 또한 화물 적재 조건이 접안속도에 상당한 영향을 미친다는 점을 강조하였다. Moriya et al.은 일반 화물선, 컨테이너선, 자동차 운반선에 대한 평균 접안속도를 Table 2.6과 같이 확립하였다.

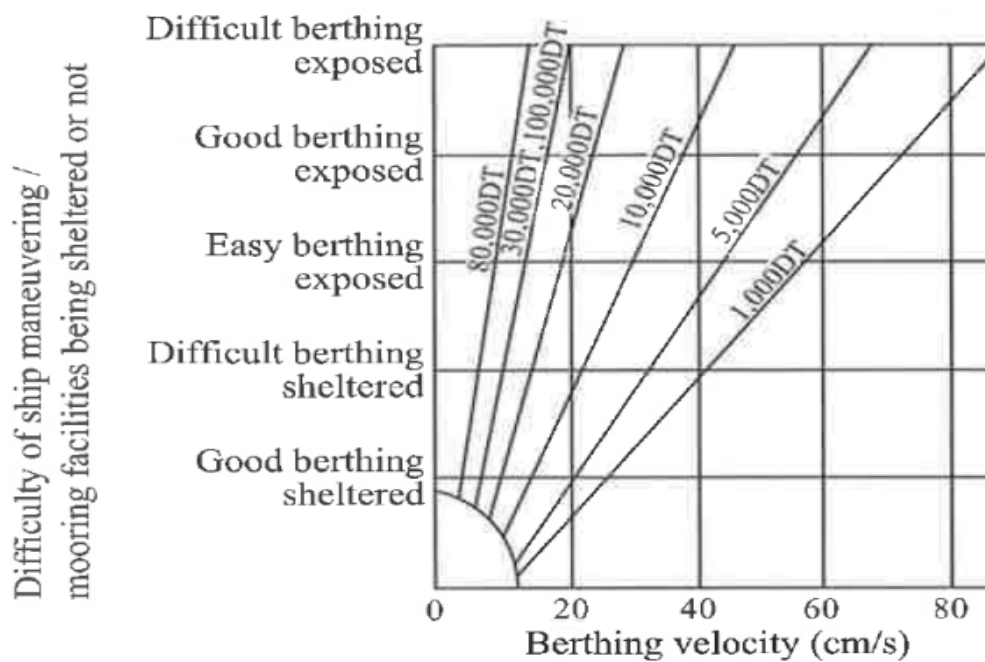


Fig. 2.5 Japanese Technical Standard chart and berthing velocity

Source : Japanese Technical Standard, 2009

Table 2.6 Berthing velocities depending on ship types

Dead Weight Tonnage (DWT)	Berthing velocity (cm/s)			
	General Cargo ships	Container ships	Pure car carriers	All ships
1,000class	8.1	-	-	8.1
5,000class	6.7	7.8	-	7.2
10,000class	5.0	7.2	4.6	5.3
15,000class	4.5	4.9	4.7	4.6
30,000class	3.9	4.1	4.4	4.1
50,000class	3.5	3.4	-	3.4
All ships	5.2	5.0	4.6	5.0

Source : Japanese Technical Standard, 2009

Ueda(1981)는 1971년부터 1973년까지 일본의 tanker 부두의 실측 접안속도 데이터를 분석하였다. 그 결과 접안속도는 Fig. 2.6과 같이 로그 정규 확률분포를 따른다는 것을 알 수 있다. 또한 Ueda et al.(1975)는 실제 선박의 접안 시 접안속도가 설계 접안속도의 약 1/3에서 1/2이어야하는 것을 주장하였다.

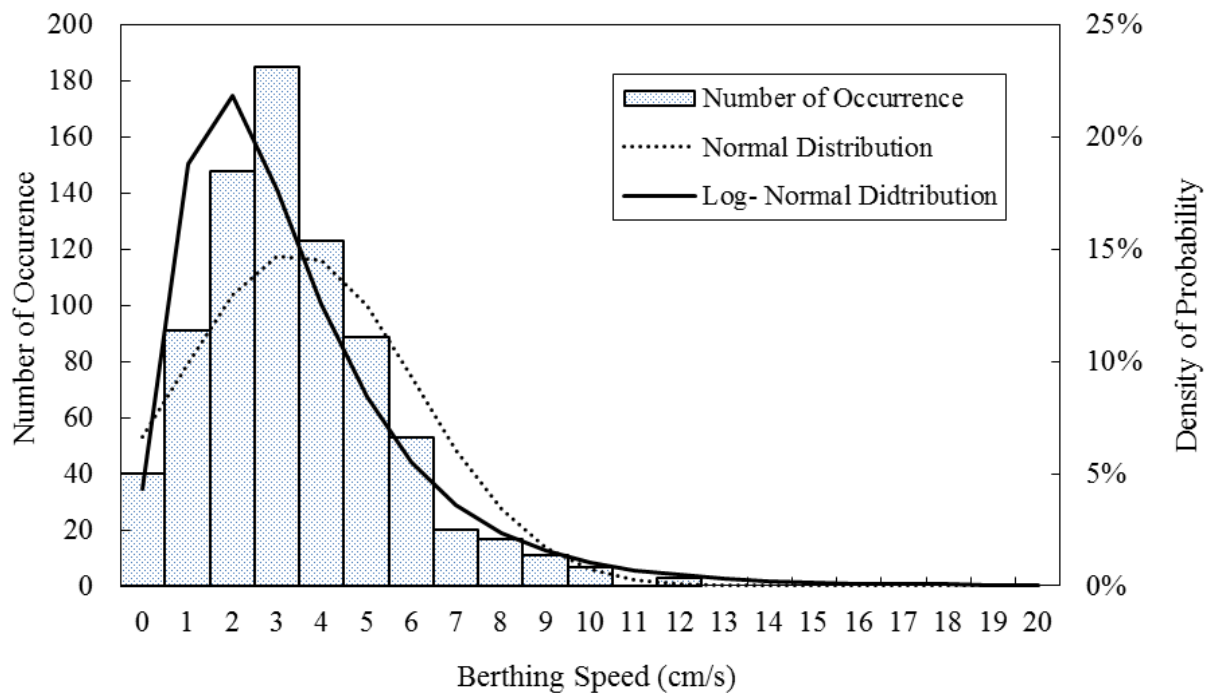


Fig. 2.6 Berthing velocity at Crude Oil Tanker Berth
(Keiyo Sea Berth in 1971~73)

Source : S. Ueda, 1981

국내의 항만 및 어항 설계기준의 경우 접안속도의 기준이 되는 자료는 일본의 과거 연구사례를 그대로 번역해 놓았다. 또한 국제적인 접안속도 관련 기준 및 연구가 활발하게 이루어지고 있는 것과는 달리 국내의 접안속도 관련 연구는 부족하기 때문에 관련 연구가 활발히 이루어질 필요가 있다.

조익순 등(2018)은 접안속도 실측 데이터를 이용해 접안속도와 접안각도를 부두별, 선박규모별로 구분하여 분석하였다. 여기서 접안속도와 접안각도는 대부분 정규분포를 따른다는 것을 확인하였다.

제 3 장 빅데이터와 머신러닝

3.1 빅데이터

빅데이터(Big Data)는 단순히 규모가 큰 데이터만을 의미하는 것은 아니다. 빅데이터의 정의는 3Vs에서 3²Vs에 이르기까지 시간에 따라 다음(Fig. 3.1)과 같은 특징을 가지며 변화하였다(Caser Wu et al., 2016).

Gartner는 빅데이터를 3V라고 해석했는데 이것이 일반적으로 알려져 있는 초기 빅데이터의 특징 3가지이다. 이는 데이터의 크기(Volume), 속도(Velocity), 다양성(Variety)으로 나누어 질 수 있다고 정의한다(Laney, 2001). 이러한 특징에 따라 빅데이터를 정의하면 빠른 속도로 생산 및 소비되는 대용량의 정형·비정형 데이터를 효율적으로 수집 및 저장하고, 다양한 분석 기법을 활용하여 현상 파악, 예측, 의사 결정 등을 지원하기 위한 기술이라고 정의 할 수 있다(박인근 등, 2019).

IBM에서는 빅데이터를 Gartner의 3V 해석에 ‘Veracity’를 추가하였는데 이는 데이터의 불확실성을 의미한다. Yuri Demchenko는 2013년에 빅데이터의 정의를 IBM의 정의와 함께 ‘Value’를 추가하여 5Vs라고 제안하였다.

Microsoft는 빅데이터의 3V 속성에 더하여 3가지 V(Variability, Veracity, Visibility)를 더 추가하여 6Vs로 확장하였다. 6가지의 빅데이터 특징을 설명하면 다음과 같다.

- 가. Volume : 데이터의 규모
- 나. Velocity : 데이터 분석 속도
- 다. Variety : 정형·비정형 데이터 등, 다양한 데이터 형태
- 라. Variability : 데이터의 복잡함 정도
- 마. Veracity : 데이터의 출처에 대한 신뢰성
- 바. Visibility : 데이터에서 정보를 도출하기 위한 시야

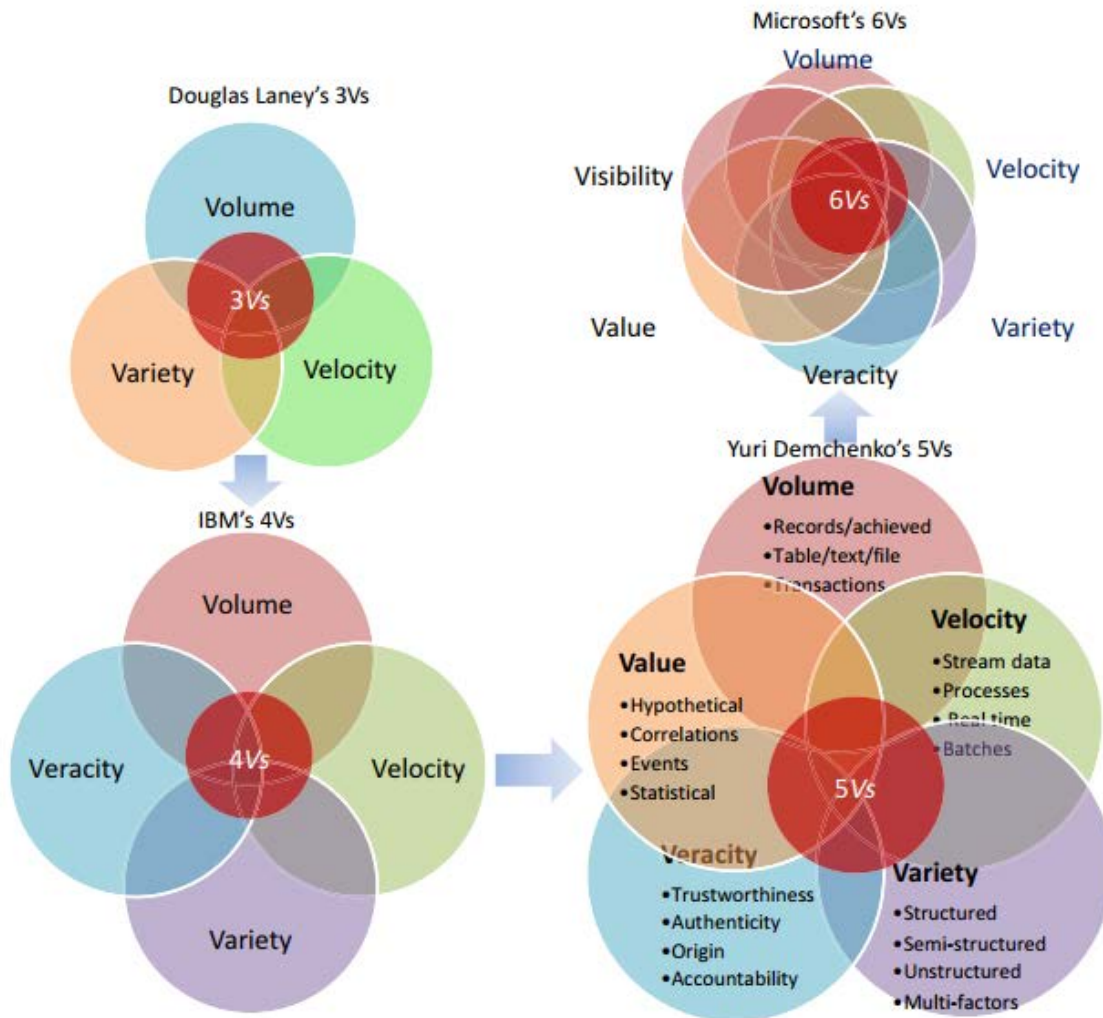


Fig. 3.1 From 3Vs, 4Vs, 5Vs and 6Vs Big Data characteristics

Source : Caser Wu et al., 2016

Caser Wu et al.(2016)은 기존의 빅데이터의 정의를 바탕으로 빅데이터의 모든 특성들이 포함된 포괄적인 빅데이터에 대한 정의를 3²Vs의 특성으로 구분하며 Fig. 3.2와 같이 나타내었다.

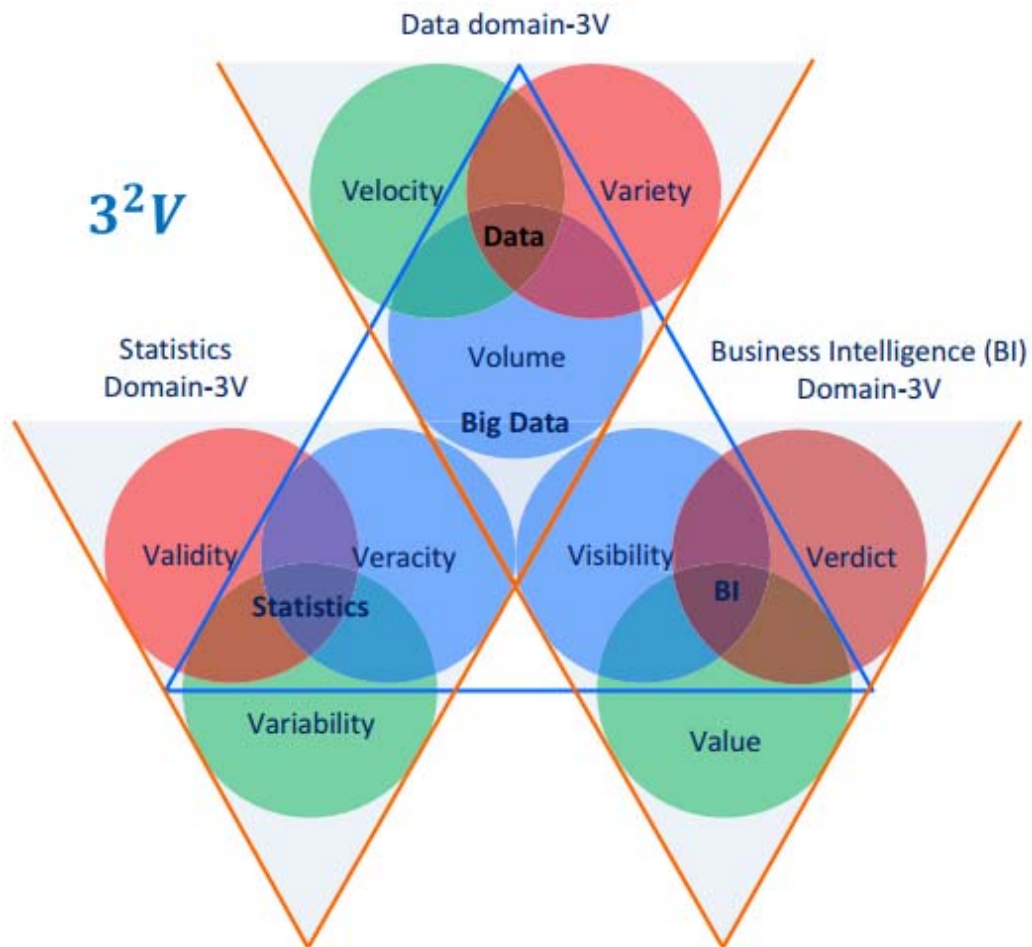


Fig. 3.2 3^2Vs Venn diagrams in hierarchical model

Source : Caser Wu et al., 2016

Fig. 3.2와 같은 3^2Vs 가 빅데이터의 의미를 나타낸다면 빅데이터 분석(Big data analysis)는 빅데이터의 실용적 의미를 나타낸다. 빅데이터 분석은 3^2Vs 의 데이터 영역, 통계영역, BI영역이 각각 Massive datasets, Machine Learning, Cloud computing으로 설명되며, Fig. 3.3는 이를 나타낸다(Caser Wu et al., 2016).

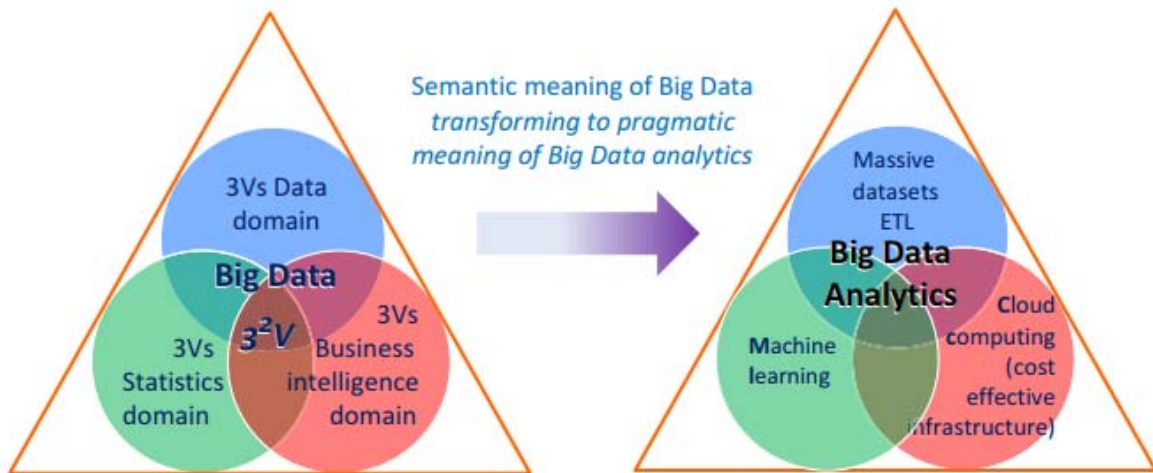


Fig. 3.3 Correlation of 3²Vs to machine learning Venn diagrams

Source : Caser Wu et al., 2016

이러한 빅데이터 분석에 의한 연구 및 활용이 활발해지면서 과거의 데이터를 통한 예측이 가능해지게 되었으며, 의사결정에 있어서도 손쉬운 해답을 찾을 수 있게 되었다. 이는 구글, 아마존 등 여러 인터넷 사이트, 기업, 공공기관 등의 사례를 이용해서도 쉽게 찾아 볼 수 있다(함유근 등, 2012). 또한, 선박 및 해운분야에서도 빅데이터 활용에 대한 필요성을 인지하여 관련 연구가 활발히 시작되고 있다.

Ornulf Jan Rodseth et al.(2016)은 해운 분야에서 선박 센서와 데이터 수집 장치, 위성 통신 등 빅데이터 수집에 대한 제약이 있으며, 새로운 프로토콜이 해운분야의 데이터 수집, 구성 과정을 단순화 할 수 있다고 제시하였다.

Mario Milicevic et al.(2018)은 해운산업은 변화하는 환경에 신속하게 적응하고, 수많은 변수를 고려하여 의사결정을 하는 복잡한 시스템이라고 언급하였다. 선박 및 해운분야도 통신, 센서 등이 발전함에 따라 데이터 축적이 크게 향상되어 빅데이터 분석을 활용한다면 의사결정에 통찰력을 얻을 수 있을 것이라고 주장하였다.

3.2 머신러닝

머신러닝(Machine Learning)의 기존의 의미는 ‘컴퓨터 또는 기계에게 프로그래밍 되지 않고 배울 수 있는 능력을 부여하는 연구분야’ 였다(Samuel, 1959). 하지만 오늘날의 머신러닝의 의미는 빅데이터 분석 기법 중 하나로써 데이터의 패턴을 자동으로 인식하고, 그 패턴을 이용해 미래에 대한 예측이나 의사결정을 하는 방법으로 정의된다(Kevin P. Murphy, 2016). 머신러닝은 다음과 같은 특징(Table 3.1)들로 그 해석을 정리할 수 있다.

Table 3.1 Interpretation of Machine Learning

Data	Information	Knowledge	Intelligence
Data mining	Information analytics	Real-time analytics	Business analysis
Data science	Information visualization	Predictive analytics	Business intelligence
Data warehouse	Information system management	Machine Learning	Artificial intelligence
Learning from data	Text analytics	Knowledge base system	Decision support system
Data smart	Text mining	Pattern recognition	Actionable intelligence
Data analytics	Web analytics	Statistical application	Business forecasting
Making sense of data	Web semantic analysis	Knowledge discovery	Business strategy
Data ingestion	Web searching	Expert systems	Business transformation

Source : Caser Wu et al., 2016

Table 3.1과 같은 특징에 따라 머신러닝은 빅데이터를 기반으로한 인공지능 (artificial intelligence, AI)의 일부라고 할 수 있다. 이는 컴퓨터 또는 기계가 복잡한 데이터 또는 문제를 입력하여 스스로 학습하게 하고 인간수준의 능력을 발휘하는 것을 뜻하기 때문이다. 이에 발전된 딥러닝(Deep Learning)은 머신러닝의 한 분야이지만 약간의 차이점이 있다. 딥러닝은 컴퓨터 또는 기계가 학습조차도 스스로 판단하여 실행하고 미래의 상황을 예측하는 분야이다. 딥러닝의 한 예로는 전 세계가 주목한 2016년 3월 딥마인드 챌린지 매치(Google Deepmind Challenge match)에서 펼쳐진 알파고(AlphaGo)와 이세돌 9단의 대국이 있다. 빅데이터, 인공지능, 머신러닝, 딥러닝의 관계도는 Fig. 3.4와 같이 설명된다(전해남, 2018).

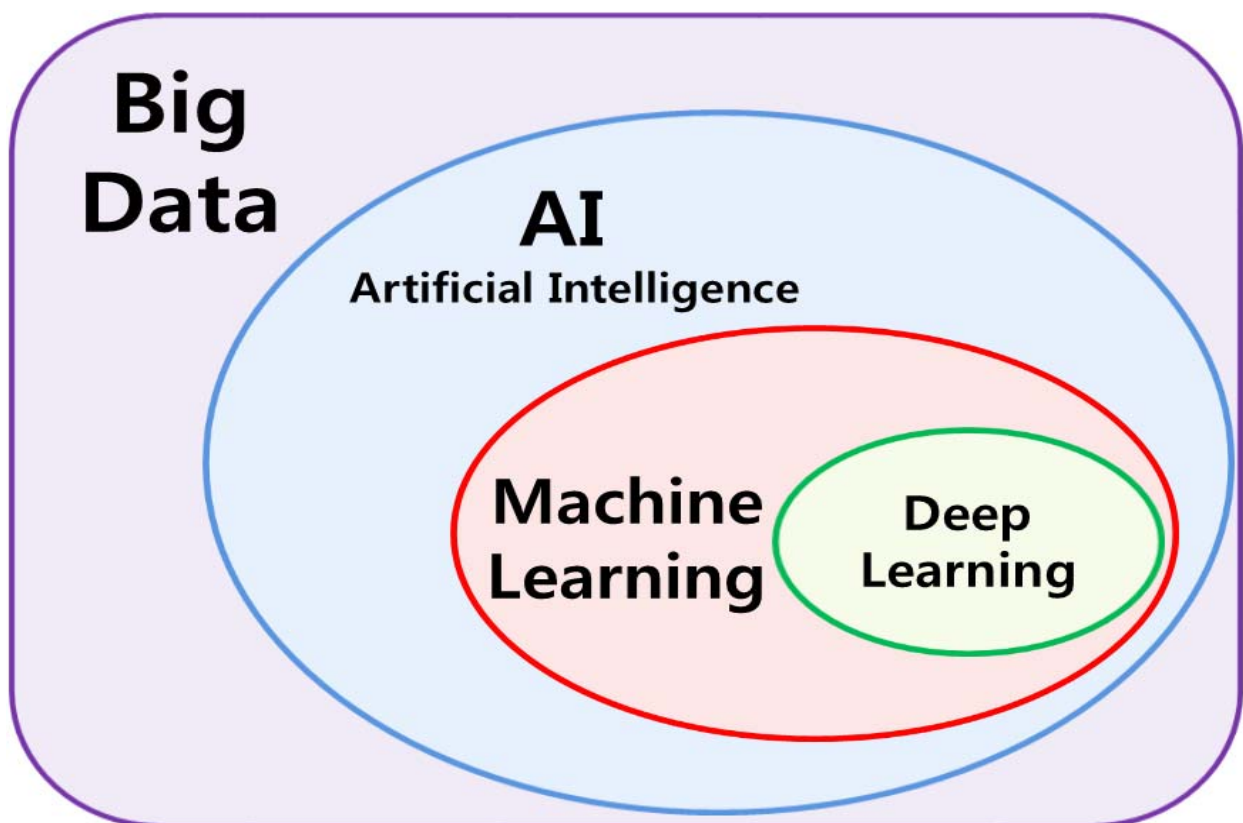


Fig. 3.4 Relationship diagram of Big Data, Machine Learning and Deep Learning

머신러닝은 학습종류에 따라 3가지로 분류하는데 지도학습(Supervised Learning)과 비지도학습(Unsupervised Learning), 강화학습(Reinforcement Learning)으로 나뉜다(Raghav Bali et al., 2017; 박인근 등, 2019).

지도학습은 주로 데이터를 통한 예측모델을 구축하는데 많이 쓰인다. 예측모델은 데이터와 머신러닝 알고리즘으로부터 추측 가능한 값들을 학습하여 얻어진다. 지도학습 알고리즘은 입력된 데이터의 특성과 예측 결과 값의 연관성 및 관계를 모델링하는 것이다. 지도학습 알고리즘은 분류와 회귀 알고리즘으로 나뉜다.

- 분류 알고리즘(Classification algorithm) : 카테고리나 레이블을 가진 데이터를 바탕으로 예측모델을 만든다. 새로운 데이터가 입력되었을 때 그 카테고리나 레이블을 분류하기 위한 예측에 사용된다.

- 회귀 알고리즘(Regression algorithm) : 데이터로부터 얻은 특성을 바탕으로 결과값을 예측한다. 기존 데이터를 통한 특성 및 결과값을 통해 모델을 만들고, 이 모델을 사용하여 새로운 데이터에 해당하는 값을 예측하는 알고리즘이다.

- 인공신경망(Artificial Neural Network, ANN) : 인간 두뇌의 학습 과정을 뉴런과 노드의 망 등을 통한 연산 과정으로 간주하여 만들어진 분류, 예측 모형이다. 데이터를 뉴런과 노드의 망을 통해 입력하여 분석하고 출력한 뒤 결과를 도출한다.

비지도학습은 입력된 데이터로부터 규칙, 패턴을 도출하여 데이터를 요약하거나 그룹화하여 결과를 해석하고 데이터를 분석하는 방법이다. 이에 대한 결과 값에는 카테고리나 레이블은 존재하지 않는다.

- 군집화 알고리즘(Clustering algorithm) : 입력 데이터로부터 도출한 특징을 이용하여 다양한 카테고리로 군집화 또는 그룹핑(Grouping)하는 알고리즘이다. 분류 알고리즘과 다른 점은 결과 값을 도출하기 전까지는 결과 레이블을 알 수 없다.

- 연관성 규칙 학습 알고리즘(Association rules algorithm) : 이 알고리즘은 데

데이터의 규칙과 패턴을 알아내고자 할 때 사용된다. 데이터 내의 각기 다른 변수와 속성을 분석하여 도출하고 그 사이의 관계를 규칙을 설명한다.

- 차원 축소 알고리즘(Dimensionality Reduction algorithm) : 차원 축소 알고리즘은 보통 데이터가 복잡하고 높은 차원을 가져 분석 및 시각화 등이 어려울 때 사용된다. 이는 데이터에 적합한 머신러닝 기법을 찾기 이전의 초기단계에도 사용된다. 크게 속성 선택(feature selection)과 속성 추출(feature extraction)으로 나눌 수 있다.

강화학습은 어떤 환경 안에서 정의된 에이전트가 현재의 상태를 인식하여 선택 가능한 행동 중 행동에 따른 보상을 얻게 하면서 최선의 행동 또는 행동순서를 선택하는 방법이다(박인근 등, 2019). 이와 같은 머신러닝의 3가지 학습방법에 따라 알고리즘의 종류를 Table 3.2과 같이 정리하였다.

Table 3.2 Machine Learning algorithm

Learning	Analysis	Algorithm
Supervised Learning	Classification	KNN (K-Nearest Neighbors) NBC (Naive Bayes Classification) Support Vector Machine Decision Tree Random Forest Cross-Validation Ensemble Learning Boosting Hidden Markov model
	Regression	Linear Regression Logistic Regression Ridge Regression Lasso Regression Jackknife Regression
	Artificial Neural Network, ANN	CNN (Convolutional Neural Network) RNN (Recurrent Neural Network) DQN (Deep Q-Network)
Unsupervised Learning	Clustering	K-means Clustering Density-based Clustering Fuzzy Clustering SOM, Self-Organization Map
	Association rules	Association rules Naive Bayes
	Dimensionality Reduction	PCA (Principle Component Analysis) ICA (Independent Component Analysis) NNMF (Non-negative matrix factorization) SVD (Singular Value Decomposition)
Reinforcement Learning	Monte Carlo Method TD-Learning, Temporal Difference Learning Policy Gradient Algorithm	

머신러닝 알고리즘 중 특히 지도학습 알고리즘이 미래에 대한 예측이 가능하고 예측을 통한 예방 및 대비가 가능하기 때문에 각 산업분야에서는 머신러닝 지도학습 알고리즘 머신러닝을 이용한 연구가 활발히 진행되고 있다.

남규민(2018)은 데이터마이닝 기법을 이용하여 기업부실 예측모형의 성과를 비교하였다. 머신러닝 지도학습 알고리즘인 로지스틱회귀(Logistic Regression), 의사결정나무(Decision Tree), 인공신경망(Artificial Neural Network, ANN)을 이용하여 부실기업과 정상기업에 대한 기업부실 예측모형의 성과를 비교하였다.

김지혜(2018)은 조선 생산 리드타임 예측을 위한 기계학습 방법론을 연구하였다. 빅데이터 기반의 머신러닝 알고리즘 연구가 조선분야의 설계분야에서는 이루어지고 있으나 생산관리 관점에서 연구가 미흡하다고 언급하였다. 조선소의 공정에 따른 생산 리드타임이라는 시간요소에 대한 개선을 위하여 머신러닝 알고리즘을 적용하여 예측모델을 구축하였다.

임상섭(2018)은 벌크선 용선에 수반되는 기간연장옵션의 가격 결정 문제와 일정기간 확보된 선박을 이용하여 다양한 대선 전략을 선택하는 문제를 다루며 머신러닝 기법을 활용하였다. 머신러닝 기법의 비교를 통해 최적의 해운의사결정문제 적용가능성을 제시하였다.

본 논문의 목적은 선박의 접안속도를 이분화해서 적정 접안속도 예측모델을 만들고 접안속도에 미치는 요인을 분석하는 것이다. 예측대상인 ‘접안속도’가 출력되는 값으로 정의되며 예측모델을 생성하기 위하여 접안속도에 영향을 미치는 요인들을 입력값으로 가진다. 따라서 본 논문에서는 머신러닝 기법 중 지도학습을 적용하며 그 중에서도 예측에 대표적인 알고리즘인 의사결정나무(Decision Tree), 랜덤포레스트(Random Forset), 로지스틱회귀(Logistic Regression), 인공신경망(Artificial Neural Network, ANN)을 이용하여 예측모델을 만들고자 한다. 예측모델을 이용하여 접안속도에 미치는 요인을 분석할 수 있다. 머신러닝 기법을 이용한 본 논문의 데이터 분석 과정은 Fig. 3.5와 같다.

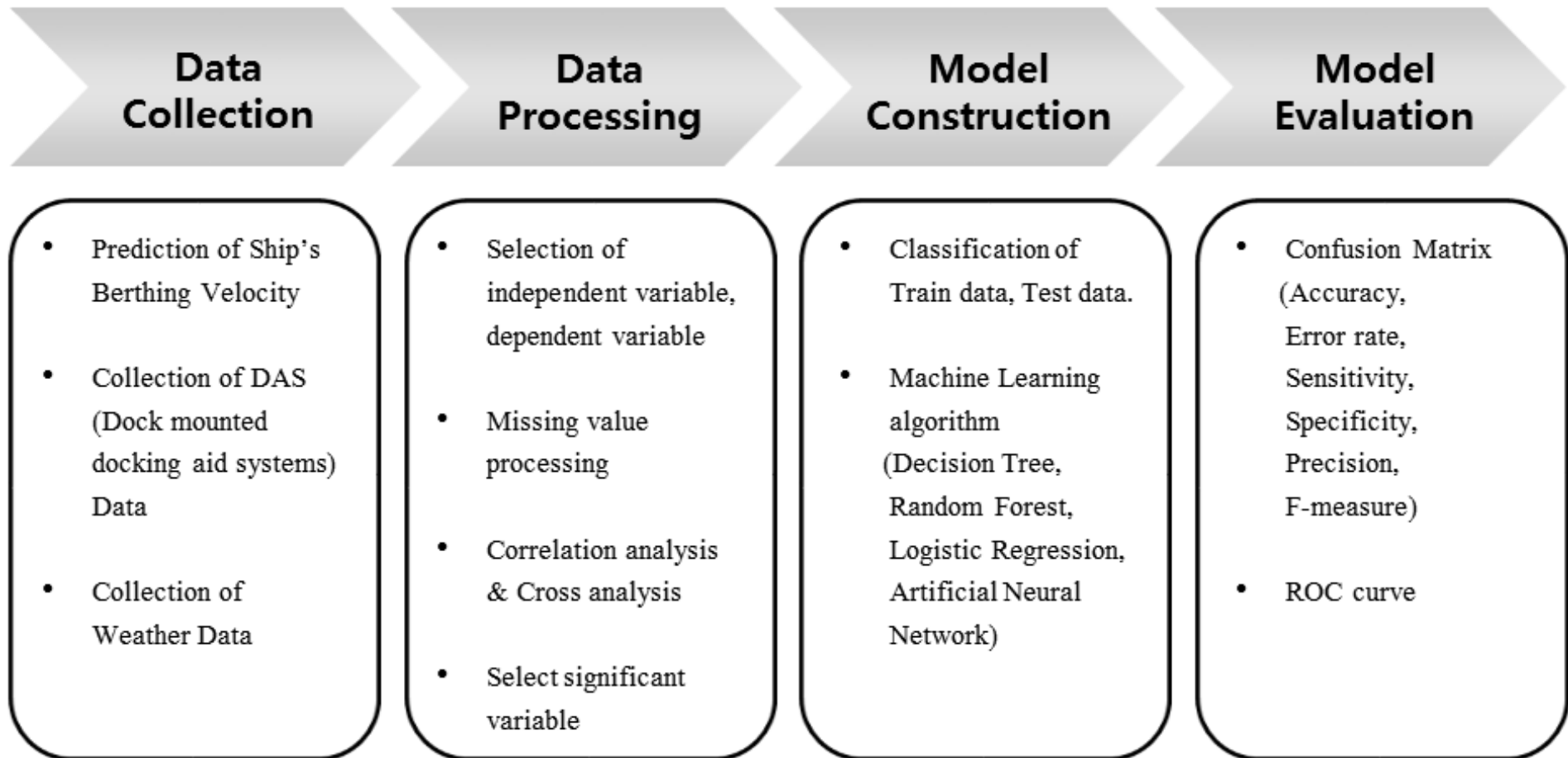


Fig. 3.5 Data analysis process

3.3 의사결정나무

의사결정나무(Decision tree)는 수집된 데이터들을 분석하여 존재하는 패턴(범주별 특성)을 속성의 조합으로 나타내는 분류(Classification) 및 예측(Prediction) 모형이다. 의사결정나무는 그 결과를 명확하고 이해하기 쉽게 설명할 수 있는 장점이 있으며 노드라고 불리는 요소들로 이루어진 나무(Tree) 형태로 만들어진다. 데이터를 각 노드에서 요구되는 범주로 분류하는 방법으로 시작하여, 데이터를 분할하는데 있어 최상의 방법을 찾기까지 가지치기(pruning)을 통하여 완성하게 된다. 의사결정나무의 일반적인 구조는 Fig 3.6과 같다(박인근 등, 2019; 김병수 등, 2018; 이정진, 2011).

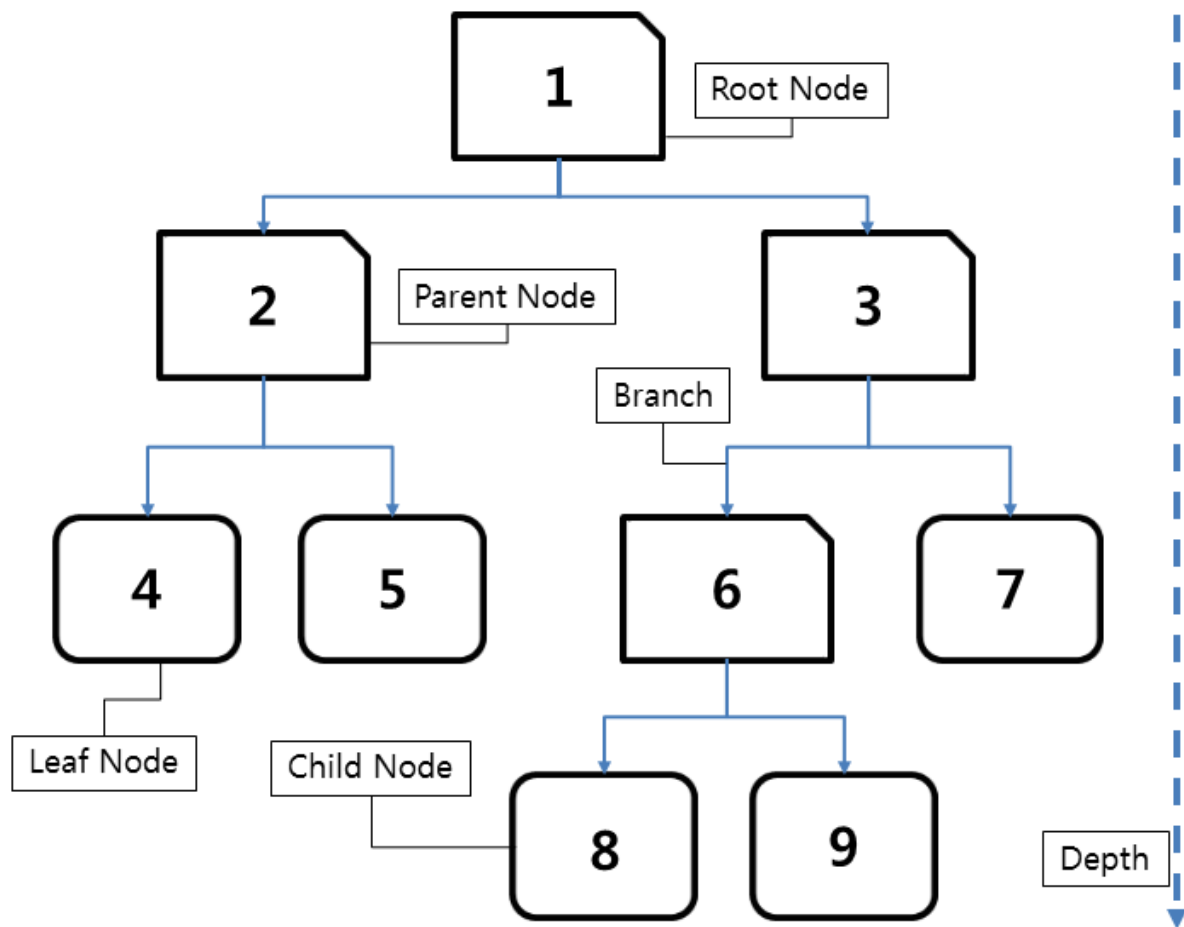


Fig. 3.6 Example of Decision Tree

의사결정나무는 Fig 3.6과 같이 노드라고 불리는 구성요소들로 이루어져 있으며, 다음과 같이 분류된다(김병수 등, 2018).

1. 뿌리노드(Root Node)는 나무가 시작되는 노드(1번)를 말한다.
2. 자식노드(Child Node)는 하나의 노드로부터 분리된 2개 이상의 노드를 말한다. 8, 9번 노드는 6번 노드의 자식노드이다.
3. 부모노드(Parent Node)는 자식노드의 상위노드를 말한다. 3번 노드는 6, 7번 노드의 부모노드이다.
4. 잎노드(Leaf Node)는 각 나무줄기의 끝에 위치하고 있는 노드이다. 4, 5, 7, 8, 9번 노드가 이에 해당한다.
5. 중간노드(Internal Node)는 나무구조의 중간에 있는 잎 노드가 아닌 노드이며 2, 3, 6번 노드가 여기에 속한다.
6. 가지(Branch)는 노드와 노드를 이어주는 역할을 한다.
7. 깊이(Depth)는 뿌리노드로부터 잎노드까지의 가지를 이루고 있는 노드 개수 중 최대값을 의미한다. 이로써 1, 3, 6, 8, 9번 노드가 하나의 가지를 이루고 있으므로 깊이는 3이 된다.

의사결정나무를 이용한 분류는 각 노드에 어떤 변수를 선택하는지가 중요한 요소이다. 목표변수의 분포를 구별하는 정도는 순수도(purity) 또는 불순도(impurity)에 의해 측정된다. 목표변수가 이산형인 경우에는 각 범주에 속하는 빈도(frequency)에 의해 분리되며, 이 때 사용되는 분리기준은 의사결정나무 알고리즘별로 카이제곱 통계량(Chi-square statistics)의 유의확률(p-value), 지니계수(Gini index), 엔트로피지수(entropy index) 등이 있다(김병수 등, 2018).

분리기준에 따라 의사결정나무의 알고리즘에는 ID3, CART(Classification and Regression tree), C4.5(C5.0) 등의 다양한 알고리즘이 존재한다. 그 각각의 특징은 Table 3.4와 같다.

Table 3.4 Basic characteristic of decision tree algorithms

Characteristic(→)	Splitting Criteria	Attribute type	Missing values	Pruning Strategy	Outlier Detection
Algorithm(↓)					
ID3	Information Gain	Handles only Categorical value	Do not handle missing values.	No pruning is done	Susceptible on outliers
CART	Towing Criteria	Handles both Categorical and Numeric value	Handle missing values	Cost- Complexity pruning is used.	Can handle outliers
C4.5 (C5.0)	Gain Ratio	Handles both Categorical and Numeric value	Handle missing values	Error Based pruning is used.	Susceptible on outliers

Source : Sonia Singh et al, 2014

본 연구에서는 의사결정나무 알고리즘 중 C5.0을 사용하였는데 C5.0은 C4.5에서 한 단계 더 발전된 알고리즘으로서 Ross Quinlan에 의해 만들어졌다. Ross Quinlan은 1986년 의사결정나무 알고리즘인 ID3를 처음으로 제안하였고 이후 1993년 발전된 개념인 C4.5를 제안하였다. ID3, C4.5, C5.0 모두 머신러닝, 인공지능 분야에서 개발되고 발전되었으나, CART를 비롯한 다른 알고리즘은 통계학분야에서 개발되고 발전되었다. 따라서 본 연구에서는 머신러닝 분야에서 활용되는 C5.0 알고리즘을 채택하여 분석을 진행하였다.

3.4 랜덤포레스트

랜덤포레스트(Random Forest)는 의사결정나무의 확장된 개념으로서 다수의 나무(tree)가 숲(forest)을 이룬 형태를 생각하면 된다. 랜덤포레스트는 많은 입력변수를 다룰 수 있으며 정확도가 높은 편이나 의사결정나무에 비해 결과 해석은 어려운 편이다. 의사결정나무는 결과 또는 성능의 변동폭이 크며 학습데이터에 따라 결정되는 나무의 형태가 크게 달라져 일반화하기 어려운 과적합(Overfitting)의 문제가 있다. 또한 의사결정나무는 계층적 접근 방식에 의한 방식으로 중간 노드에서 오류 발생 시 다음 노드까지 그 오류가 이어지는 문제가 발생한다. 이런 의사결정나무의 한계점으로 인하여 랜덤포레스트가 등장하게 되었다(박인근 등, 2019).

랜덤포레스트는 배깅(bagging)기법과 부스팅(boosting), 임의 노드 최적화(randomized node optimization) 개념을 포함하고 있다. 배깅(bagging)이란, 주어진 데이터를 통해 여러 개의 붓스트랩(bootstrap)을 생성하여 각 붓스트랩 별 모델링한 후 결합하여 최종 예측 모델을 산출해내는 방법을 말한다. 여기서 붓스트랩이란 단순 복원 임의 추출법을 통해 기존의 데이터로부터 추출한 크기가 동일한 여러 개의 표본자료를 말한다. 배깅 기법에 대한 개념도는 Fig 3.7와 같다.

부스팅(boosting) 알고리즘은 랜덤포레스트 기법 중 하나로 잘못 분류된 개체를 더욱 잘 분류하기 위해서 새로운 분류 규칙을 만드는 단계를 반복하는 것을 말한다. 부스팅은 예측모델의 정확도를 향상하는 방법이며 예측변수를 차례로 생성하는 특징을 가지고 있다(박인근 등, 2019).

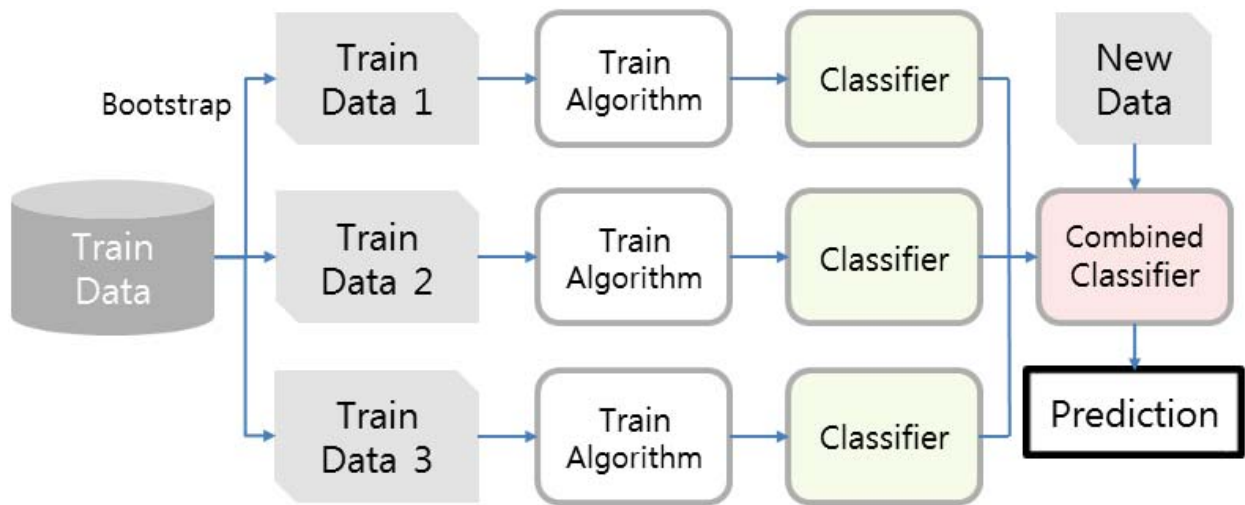


Fig. 3.7 Bagging Algorithm

임의 노드 최적화(randomized node optimization)은 분석에 사용되는 변수를 임의적으로 추출하는 것을 말한다. 모델을 훈련하는 단계서 목적 함수를 최대 로 만드는 노드분할 함수의 매개변수 세타의 최적값을 구하는 과정이다(박인근 등, 2019). 중요 매개변수로는 전체 포레스트를 몇 개의 트리로 구성할지를 결정하는 포레스트의 크기, 최대 허용 깊이, 임의성 등이 있다.

랜덤포레스트는 위의 특성을 가지고 Fig. 3.8과 같은 알고리즘을 가진다. 첫 번째 단계로 Train data set에서 복원 추출에 의해 여러 개의 부스트랩 데이터를 생성한다. 그 다음 단계로 부스트랩 데이터를 의사결정나무 알고리즘에 적용한다. 그리고 전체변수 중에서 m 개의 변수를 임의로 선택하여 나무를 성장시킨다($mtry$). 이 때 개별 모형들 사이의 상관관계가 줄어들기 때문에 모형 성능의 변동이 감소하는 효과가 있다. 마지막 단계로 전 단계에서 만든 의사결정나무 모형을 하나의 모형으로 결합하면 랜덤포레스트 알고리즘이 완성된다.

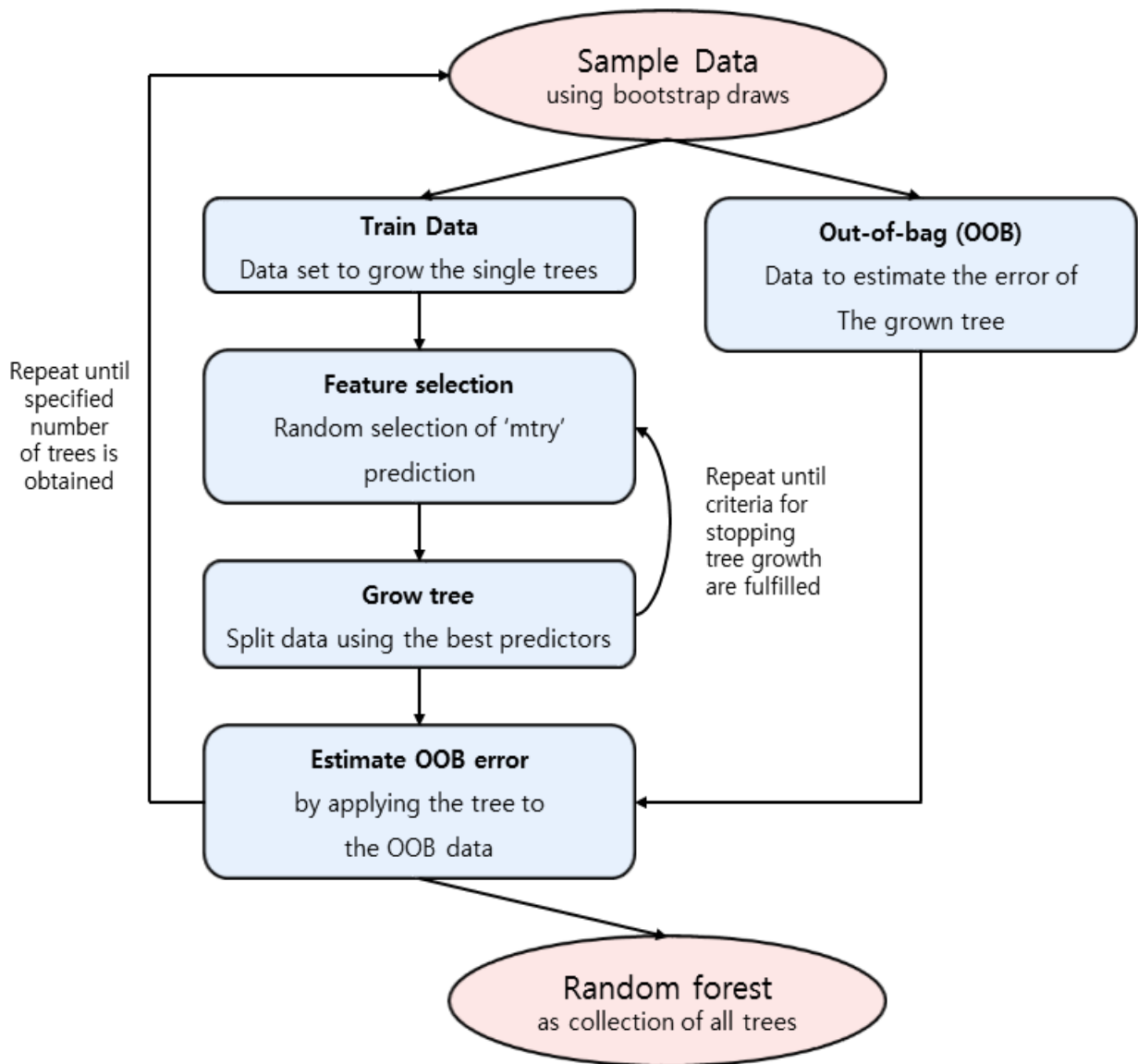


Fig. 3.8 Flow of Random Forest algorithm

3.5 로지스틱회귀

로지스틱회귀(Logistic Regression) 모형은 일반적인 회귀분석과는 다르게 종속변수가 범주형 또는 연속형 모두 분석 모형으로 사용할 수 있다. 독립변수가 선형임을 이용하여 사건의 발생 가능성을 예측하는 방식으로서 일반적인 회귀분석과 유사하지만 종속변수가 두 집단으로 나누어진 이분변수일 때 확률에 대한 로짓(logit)변환을 하여 분석하는 방법이다. 종속변수가 2개일 때는 이분형 로지스틱, 3개 이상일 때는 다항형 로지스틱 회귀분석이라고 한다.

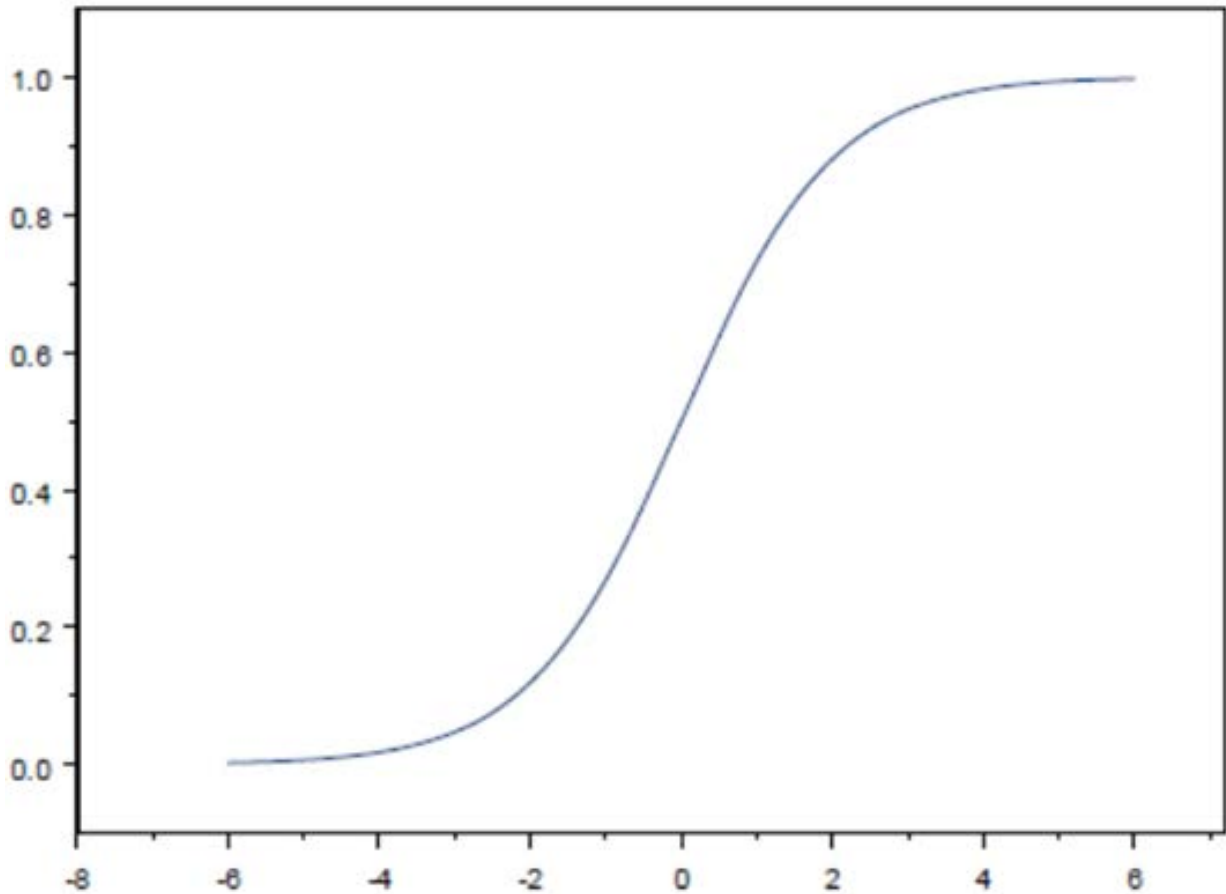


Fig. 3.9 Example of Logistic curve

로지스틱 회귀분석의 등식은 식 3.1과 같다.

$$\frac{\pi}{1-\pi} = e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i} \quad (3.1)$$

식 4.1에서 앞부분은 승산비(Odd Ratio, OR)로서 사건이 발생할 확률과 사건이 발생하지 않을 확률의 비율을 나타낸다. 등식양변에 자연로그를 취하면 로짓모형으로 변환되고 독립변수에 대한 선형관계(식 3.2)가 성립된다(조국연, 2018).

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i \quad (3.2)$$

로지스틱 회귀분석의 목적은 추정된 로짓모형을 이용하여 종속변수의 사후 확률을 구하는 데 이용된다. 위의 모형식을 $P(Y=1|X_1, X_2, \dots, X_i)$ 에 대하여 정리하면 식 3.3과 같다.

$$P(Y=1|X_1, X_2, \dots, X_i) = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i)} \quad (3.3)$$

위의 모형식으로 추정된 사후확률은 0과 1사이의 값을 가지게 되므로 기준값을 정하여 0과 1로 분류하게 된다(김병수 등, 2018; 이정진, 2011). 본 논문에서는 0이 접안속도 10cm/s 미만, 1이 접안속도 10cm/s 이상으로서 예측을 하도록 모델을 구축하였다.

3.6 인공신경망

인공신경망(ANN, Artificial Neural Network)은 빅데이터 분석 알고리즘 중 현재 가장 활발히 연구 및 활용되고 있는 분야로서 인간 두뇌의 학습 과정을 뉴런과 노드의 망 등을 통한 연산 과정으로 간주하여 만들어진 분류, 예측 모형이다. 인간이 경험으로 학습하는 두뇌의 신경망 활동, 즉 생물학적으로 뇌가 감각 입력의 자극에 반응하는 원리를 그대로 활용하여 설계된 기법으로 단순한 선형 패턴뿐 아니라 복잡한 비선형 패턴까지 학습할 수 있어 높은 예측성과를 보이는 것이 특징이다(박인근 등, 2019).

인공신경망의 구성 요소는 노드, 가중치, 활성화함수, 입력층, 은닉층, 출력층으로 구성되어있다(Fig. 3.10). 노드는 신경계의 뉴런 역할로서, 가중치와 입력값의 곱으로 활성화함수를 통해 다음 노드에 전달해주는 역할이다. 입력층은 학습을 위한 기초 데이터 입력 계층이며, 은닉층은 데이터를 처리하는 구간이며 입력층과 출력층 사이에 존재한다. 출력층은 학습을 통해 도출된 결과를 출력하는 계층이다.

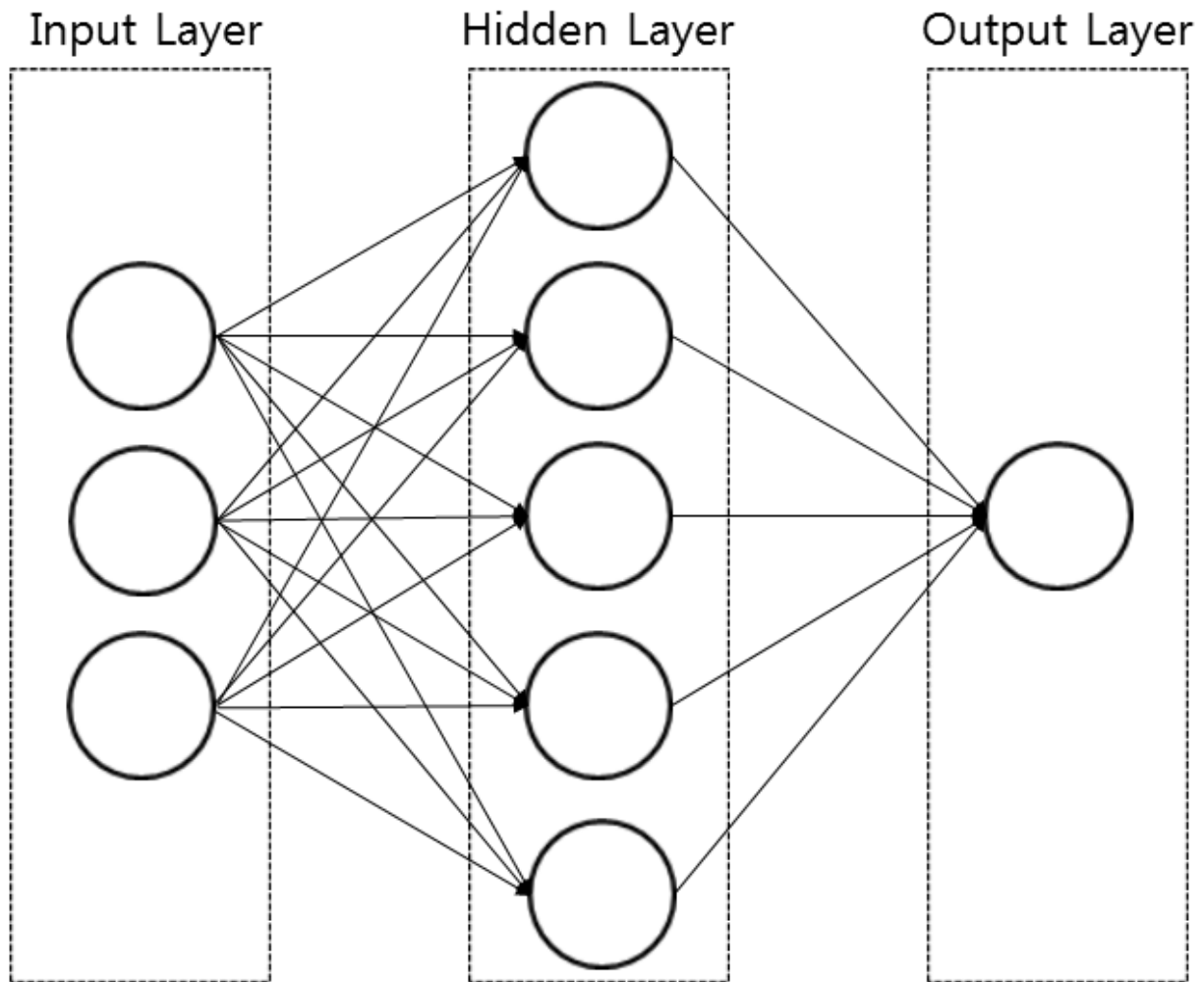


Fig. 3.10 Structure of ANN

인공신경망 모형은 다음과 같은 특성을 가진다(이정진, 2011).

- 실제 데이터에서 신경망 모형은 다른 모형보다 분류에 있어 우수한 결과를 보인다. 특히 변수의 수가 많고 입력변수와 출력변수가 복잡한 비선형함수 형태를 가질 때 유용하다.
- 인공신경망 모형은 해석이 어렵다. 이는 모형의 수정이 어렵다는 단점을 가진다.

- 다층신경망은 비선형분류함수를 추정하는 것으로서 적어도 하나 이상의 은닉층이 필요하다. 적절한 은닉층의 수와 은닉노드 수의 결정이 모형의 과적합을 피하기 위해 매우 중요하다. 인공신경망 모형은 훈련용 데이터에서는 만족스러운 결과를 보이지만, 실제 적용에서는 분류가 정확하지 않은 경우가 있는데 이는 모형의 과적합이 주요 원인이다.

- 인공신경망은 학습 데이터에 잡음이 있더라도 민감한 반응을 보이지 않는다. 따라서 입력정보의 오류를 파악하기가 쉽지 않다.

- 인공신경망의 훈련과정은 은닉층의 노드의 수가 많으면 시간이 오래 걸린다. 하지만 훈련이 끝난 후에 검증용 데이터에 대해서는 신속하게 분류할 수 있다.

제 4 장 실측 데이터 분석

4.1 데이터 수집 및 선정

4.1.1 데이터 수집

본 연구에서는 DAS가 설치되어있는 국내항 중 Jetty 1~3으로 구분된 탱커 터미널 부두에서 접안속도 실측 데이터를 수집하였다.

DAS는 선박이 부두에 접안 시 접근하는 거리와 속도를 감시 하는 시스템으로 대부분의 탱커 터미널에 설치되어 사용되고 있으며 일반적으로 Laser를 이용하여 측정한다. 보통 선석 당 2개의 Laser sensor로 측정하며, 안벽에서 최대 200m 이격된 선박 접안속도를 1cm/sec 기준으로 측정하고 대형 전광판에 표시하여 과거 도선사가 눈으로 확인하며 접안 하는 방식과 다르게 접안속도를 식별이 가능하도록 설계한다(이상원, 2019).

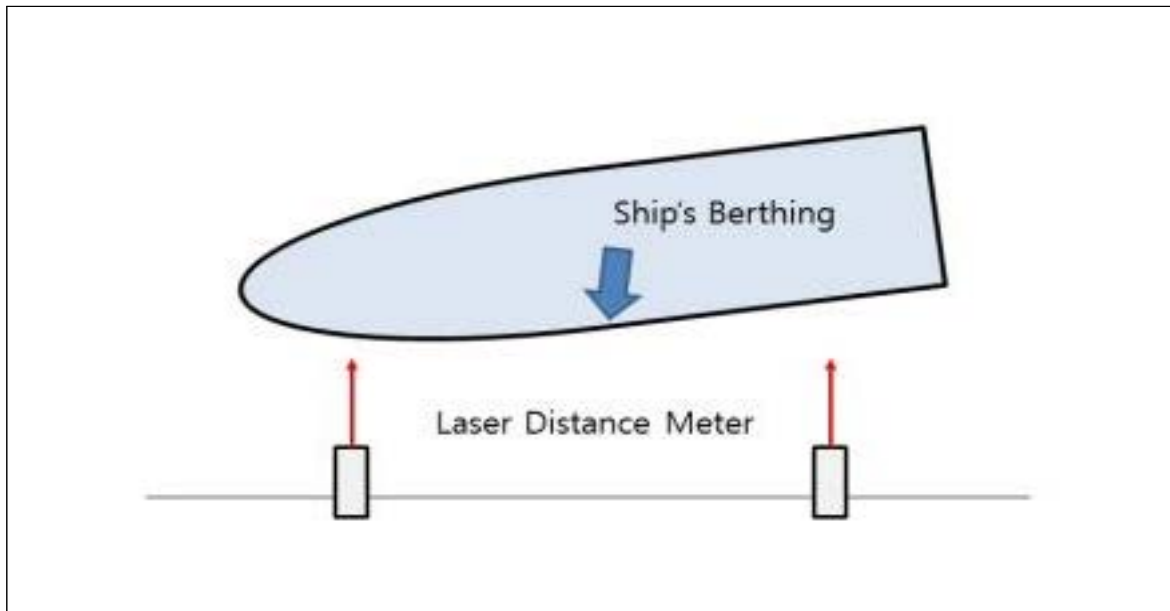


Fig. 4.1 Arrangement of Dock mounted docking aid system

‘선박의 입항 및 출항 등에 관한 법률’에서는 2017년 위험물을 운송하는 총톤수 5만 톤 이상의 선박이 접안하는 돌핀 계류시설의 운영자는 해당 선박이 안전하게 접안하여 하역할 수 있도록 해양수산부령으로 정하는 안전장비를 갖추도록 조항을 신설하였으며, 국내 항설기준에서도 일정 규모 이상의 위험물을 취급하는 돌핀부두에 접안속도계 설치를 권고하고 있다.



Fig. 4.2 SmartDock system

Source : Trelleborg(2010)

국내의 탱커 터미널의 DAS를 통해 수집된 데이터는 2017년 3월부터 2018년 7월까지 약 17개월 간 대상 부두에 접안한 선박에서 측정된 데이터를 취합하여 분석하였다. 선박의 규모는 약 10,000 DWT에서 320,000 DWT까지 다양하며 탱커 터미널이므로 선종은 탱커에 한정되어있다. 또한 접안속도는 부두의 방충재에 최초 접촉할 당시의 속도를 기준으로 하였다(Table 4.1).

Table 4.1 Particulars of measured data

Periods	March 2017 ~ July 2018
The subject of terminal	One of Tanker terminal in Republic of Korea, Jetty 1~3
Type of ship	Tanker (10,000 DWT ~ 320,000 DWT)
Method of measurement	Dock mounted docking aid system

연구를 위해 실측데이터를 측정한 부두의 특성은 Table 4.2과 같다.

Table 4.2 Berth particular and operation regulations

	Jetty 1	Jetty 2	Jetty 3
Depth	17.0 m	18.0 m	19.5 m
Capacity	80,000 DWT	120,000 DWT	320,000 DWT
Max LOA	295 m	321 m	382 m
Berthing velocity (Designed)	12 cm/s	15 cm/s	15 cm/s
Berthing velocity (Operated)	Safety : 5cm/s Warning : 6~10 cm/s Dangerous : 11~15 cm/s		
Berthing angle	Max 6 deg. within 30 m Max 3 deg. at touching		

접안속도 관련 실측데이터 이외에 수집한 날씨(바람) 데이터는 기상청(Korea Meteorological Administration)에 있는 해당 지역의 최대 풍속 기록을 참조하였다. 선박의 접안 시 시간에 해당하는 풍속을 수집하여 분석하는 것이 정확하나, 실측데이터 수집 시 선박의 접안 시간은 기록되지 않아 해당 접안일의 최대 풍속을 수집하였다.

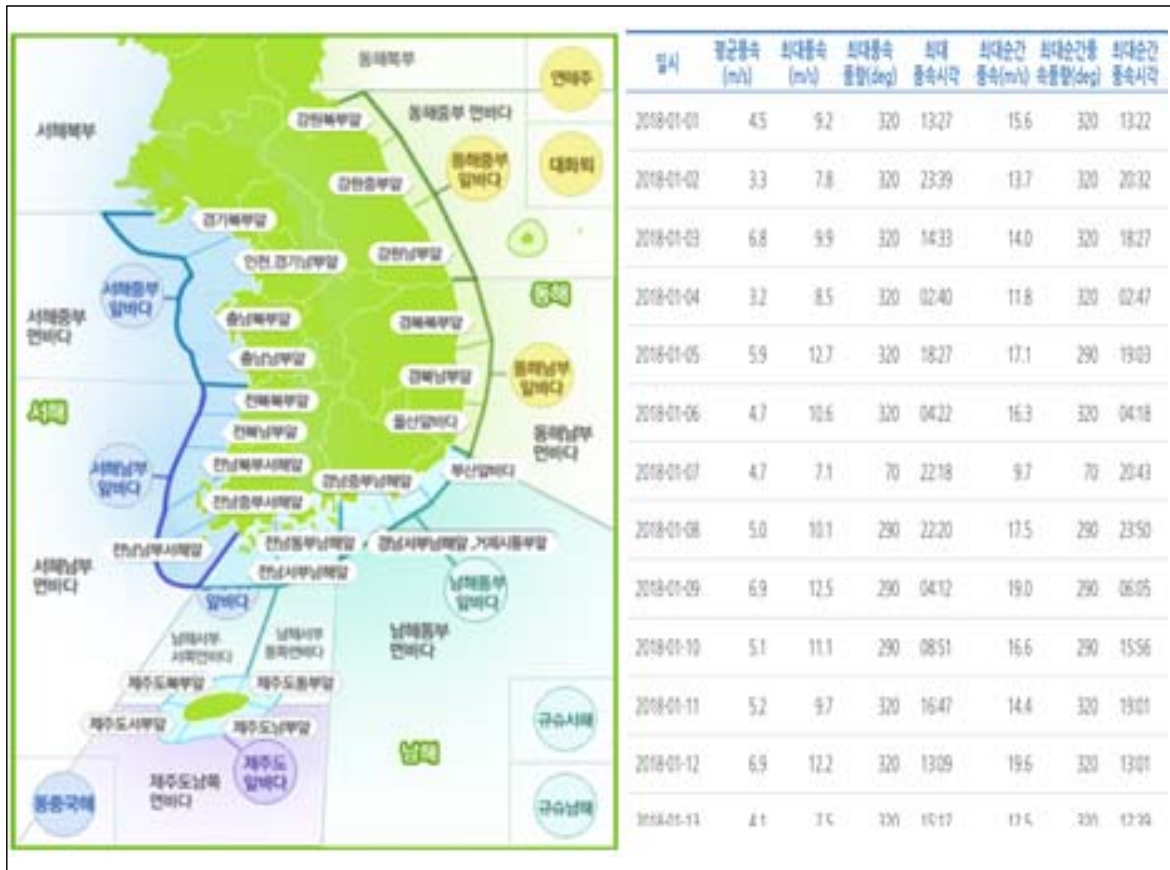


Fig. 4.3 Weather(wind) data

Source : Korea Meteorological Administration

4.1.2 데이터 선정

선박 접안 시 DAS에서 수집된 실측데이터와 부두에서 제공된 자료에는 Table 4.3과 같은 정보가 수집된다.

Table 4.3 DAS data information

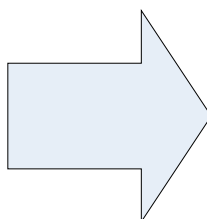
Information
Jetty No.
Vessel Name
Date
In & Out
DWT
LOA
Breath Extreme
Max Direction
Max Fender Touch(cm/s)
North Fender Deflection
South Fender Deflection
Max Angle
Pilot (main)
Pilot (Assist)
Draft(m)
Cargo Operation
Tug

본 연구에서 활용될 독립변수의 선정을 위해 DAS에서 제공되는 정보에서 PIANC WG33과 WG145에서 언급한 접안속도에 영향을 미치는 요소를 고려하여 데이터를 선별하였다(Table 4.4).

- Jetty No. : Jetty Number는 3개의 Jetty 중 선박이 접안한 선석을 나타낸 것으로 WG33의 Berth appearance와 WG145의 Closed/Open quay/jetty에 해당한다. 따라서 접안속도에 영향을 미치는 요소로 독립변수로 선택하였다.

Table 4.4 Selection of Variables

Category	Factor (WG33)	Factor (WG145)	DAS & Weather data
Environment	Prevailing physical boundary conditions	Current	Wind (Based on Korea Meteorological Administration)
		Effect waves	
		Effect wind	
	Berth appearance	Closed/Open quay/jetty	Jetty No.
Ease of navigation	-	-	
-	<u>Berthing aids</u>	DAS	
Vessel	Type of vessel	Vessel type	DWT, STATE (IN & OUT)
	Size of vessel		
	Type of cargo		
	Windage area of vessel		
Human	Method of berthing	Berthing manoeuvre	MaxAngle
	Use of tugs	Equipment / tugs	Pilot
	Human factor	Human factor	



- STATE : DAS 데이터의 IN & OUT 항목을 바탕으로 선박의 Laden, Ballast 상태를 독립변수로 채택하였다. 이 변수는 DWT와 같이 PIANC의 Vessel Type에 해당하는 변수이며, 화물의 선적 유무를 나타낸다. Laden은 선박의 만재상태이며, Ballast는 경하상태이다.

- Pilot : DAS 데이터의 Pilot(main)과 Pilot(assist) 중 메인 도선사를 바탕으로 도선사의 급수를 독립변수로 지정하였다. 이는 PIANC의 Human 항목에 해당하는 요인으로 도선사의 경험이 접안속도에 영향을 미친다는 가정하여 채택하였다. 1급도선사부터 4급도선사까지 4개의 변수로 선정하였다.

- DWT : PIANC의 Vessel Type에 해당하는 변수로서 선박의 재화중량톤수를 독립변수로 설정하였다. 이는 접안속도 관련 기준과 선행연구에서도 접안속도에 가장 영향을 미치는 요소에 해당한다.

- MaxAngle : 선박이 방충재(Fender)에 접안할 때의 접안각도를 나타낸 것이다. PIANC의 Method of berthing, Berthing manoeuvre에 해당하는 변수이다. 이 독립변수는 접안각도가 접안속도에 미치는 영향이 있는지를 가정하여 채택하였다.

- Wind : 선박이 접안한 날짜를 바탕으로 기상청을 활용하여 해당 부두의 위치의 Wind 데이터를 참조하여 변수를 설정하였다. Weather 데이터 중 Current와 Wave에 해당하는 데이터는 참조하지 않았는데, 해당 부두의 위치가 Current와 Wave의 영향이 덜하기 때문에 제외하였다. Wind 데이터는 해당 날짜에 해당하는 최대 풍속을 변수로 선정하였다. 이는 WG33과 WG145의 Weather 데이터에 해당한다.

- Berthing Velocity : 본 연구에서 분석하고자하는 종속변수이다. DAS는 선수미 방향 2개 지점에서 측정을 하는데 선박이 방충재에 접촉하는 시점의 속도를 측정한다. 이때의 선수미 방향의 2개의 접안속도 중 최댓값을 최초 접촉한 접안속도라 가정하였다.
- Berthing Aids : 본 연구에서는 DAS를 이용하여 실측데이터 수집을 하였다. 이는 해당 탱커부두의 모든 선박이 접안 시 DAS를 통해 접안속도를 Monitoring하며 접안했음을 의미한다. 따라서 Berthing Aids를 분석을 위한 변수로 설정을 하지는 않았다.

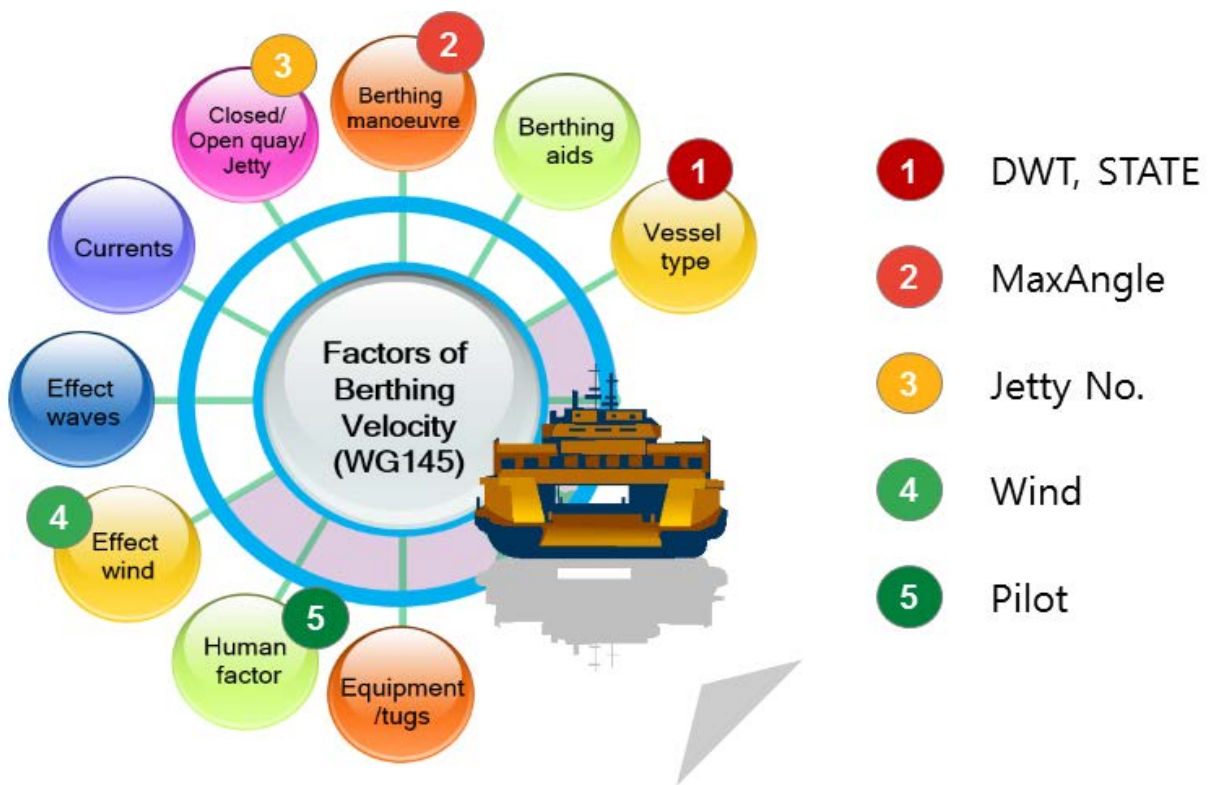


Fig. 4.4 Selection of Variables in PIANC WG145

PIANC에서 언급한 접안속도에 영향을 미치는 요인 중 예인선(Tug)의 경우에

는 실측데이터 상 Tug name만 기입되어있어 변수설정에서 제외하였다. 따라서 PIANC WG33과 WG145의 접안속도 에 영향을 미치는 요인에 의거하여 선정한 독립변수는 Jetty No., STATE, Pilot, DWT, MaxAngle, Wind로 설정하였다(Fig. 4.4). 6개의 종속변수 중 Jetty No., STATE, Pilot은 항목이 정해져 있는 범주형 변수(Categorical Variables)이며, DWT, MaxAngle, Wind의 경우에는 숫자형 변수, 즉 연속형 변수(Continuous Variables)이다.

독립변수 설정 시, STATE와 Pilot 변수 중 29개의 결측값이 존재하였다. 변수의 결측값 처리를 위하여 몬테카를로 방법(Monte Carlo Method)를 이용하여 결측값을 처리하였다. 몬테카를로 방법은 변수의 분포로부터 표본을 생성하고, 주어진 표본의 경험적 분포로부터 해당 결측값의 근사치를 계산하는 방법을 말한다. 카지노로 유명한 유럽의 한 도시이름에서 유래되었으며, 통계 물리학에서 처음으로 개발되었으나 현재는 통계학과 머신러닝 분야에서도 활발히 사용되고 있다(Kevin P. Murphy, 2016).

종속변수의 경우 데이터 분석을 위해 연속형 변수, 범주형 변수 2가지로 설정하였다. 이는 각각의 변수 별 형태에 따른 분석이 필요하며, 머신러닝 기법 적용에는 이분화된 변수 설정이 필요하기 때문이다. 데이터를 수집한 부두의 규정에 따르면 접안속도 10cm/s까지를 위험 속도로 정하며, 그 이상의 접안속도는 매우 위험한 접안속도로 정하고 있다(Table 4.5). 따라서 종속변수인 접안속도 데이터를 10cm/s 이상과 10cm/s 미만의 범주형 변수로 설정하였다.

Table 4.5 Terminal approaching Regulation

Approaching Speed	1) 5cm/sec or less : allowable speed 2) 6~10cm/sec : warning speed 3) Over 11~15cm/sec : critical (the master & pilot will be black-listed and held accountable if in case of fender damage)
--------------------------	--

독립변수, 종속변수 설정 완료 후 분석을 실시하기 전, 데이터를 코딩 하였다. 이는 분석이 용이하고 결과 해석을 간략하게하기 위함이다. 독립변수, 종속변수의 상세사항과 변수별 코딩한 값을 정리하면 Table 4.6과 같다.

Table 4.6 Variables and descriptions

(N=206)

Variable	Type	Description	Mean
Jetty No.	Categorical (Nominal scale)	1 = Jetty 1 2 = Jetty 2 3 = Jetty 3	1.879
STATE	Categorical (Nominal scale)	1, if Ballast; 0, otherwise	0.597
Pilot	Categorical (Ordinal scale)	1 = 1 st Class 2 = 2 nd Class 3 = 3 rd Class 4 = 4 th Class	1.359
DWT	Continuous (Interval scale)	Deadweight tonnages	70088.762
MaxAngle	Continuous (Interval scale)	Ship's berthing Angle (degrees)	7.519
Wind	Continuous (Interval scale)	Max wind speed (m/s)	7.883
Berthing Velocity	Continuous (Interval scale)	Max ship's Berthing Velocity (cm/s)	7.858
	Categorical (Nominal scale)	1, if Under10; 0, otherwise	0.242

4.2 기초 데이터 분석

2017년 3월부터 2018년 7월까지 약 17개월간 해당부두의 DAS에서 수집된 206개의 실측데이터를 바탕으로 종속변수, 독립변수를 선정하였다. 이 변수는 형태에 따라 범주형 변수, 연속형 변수로 분류된다. 범주형 변수는 문자로 구분된 변수이며, 연속형 변수는 숫자형 변수이기 때문에 변수를 구분하여 기초적인 분석을 통해 데이터를 해석하고자 한다. 기초 데이터 분석에는 Origin 9.0와 IBM SPSS Statistics version 25 프로그램을 이용하였다.

4.2.1 범주형 변수

범주형 변수(Categorical Variable)는 묶음을 표시하는 변수이다. 예를 들어 성별, 국적, 직업 등 Category로 나눌 수 있는 변수를 뜻한다. 범주형 변수의 분석을 위해 빈도분석을 실시하였다.

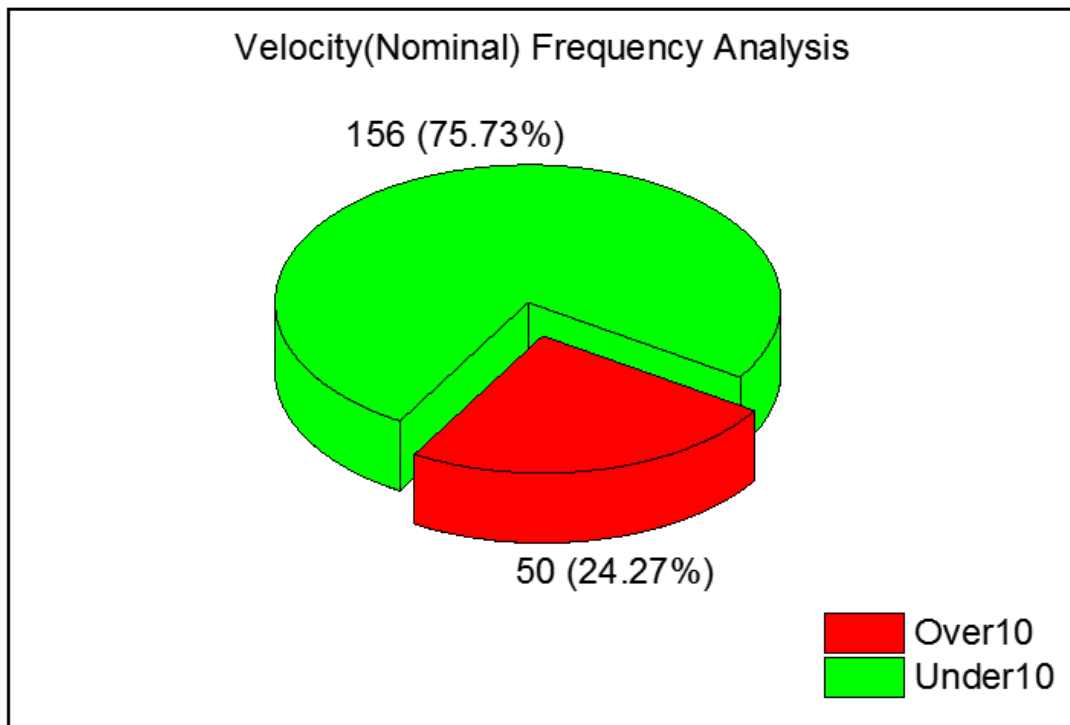


Fig. 4.5 Berthing Velocity Frequency Analysis

먼저 종속변수인 접안속도를 10cm/s 이상(Over10), 10cm/s 미만(Under10)으로 분리한 범주형 변수로 빈도분석을 진행한 결과 Under10(156회)과 Over10(50회)이 약 3대1 비율로 나타났다(Fig. 4.5). 이는 터미널 규정상 접안속도 10cm/s 이상인 경우 매우 위험(Critical)한 접안속도로 언급하였음에도 접안속도 10cm/s 이상으로 접안하는 경우가 다수 존재하였다는 것을 의미한다. 하지만 부두 설계 접안에너지에 의한 권고 접안속도를 준수하여야 부두 및 선박 안전을 확보할 수 있으므로 접안속도 10cm/s를 넘지 않도록 사전에 대비하여야 한다.

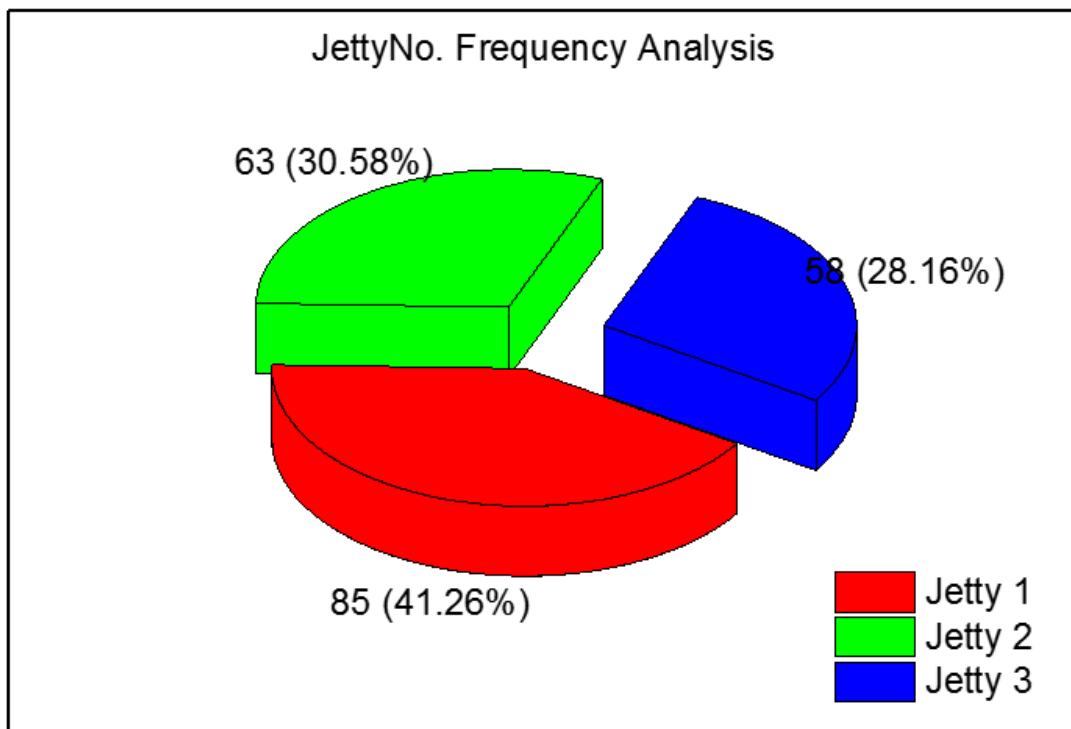


Fig. 4.6 Jetty No. Frequency Analysis

독립변수 Jetty No.의 경우 Jetty 1(85회), Jetty 2(63회), Jetty 3(58회)이 거의 비슷한 비율로 선박이 접안함을 알 수 있다(Fig. 4.6). STATE의 경우에는 Laden 상태(123회)와 Ballast 상태(83회)가 약 6대4의 비율로 나뉜다(Fig. 4.7). Pilot은 1급 도선사(166회)의 비율이 80%로 압도적으로 많은 것을 볼 수 있는데 해당 항만에서는 선박이 접안 시 대부분 경험이 많은 도선사가 승선하여 접안함을 알 수 있다(Fig. 4.8). 하지만 독립변수별 빈도분석 결과로는 특징을 찾을 수 없어 접안속도 10cm/s 이상인 경우와 10cm/s 이하인 경우로 나누어서 교차분석을 실시하였다.

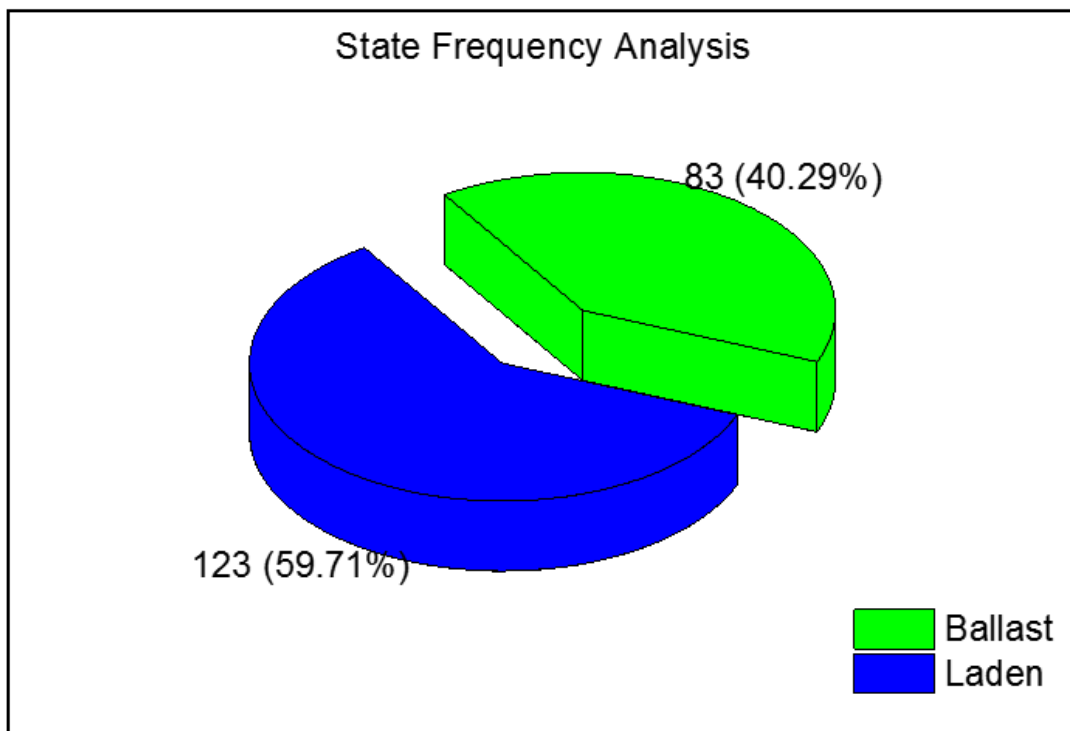


Fig. 4.7 STATE Frequency Analysis

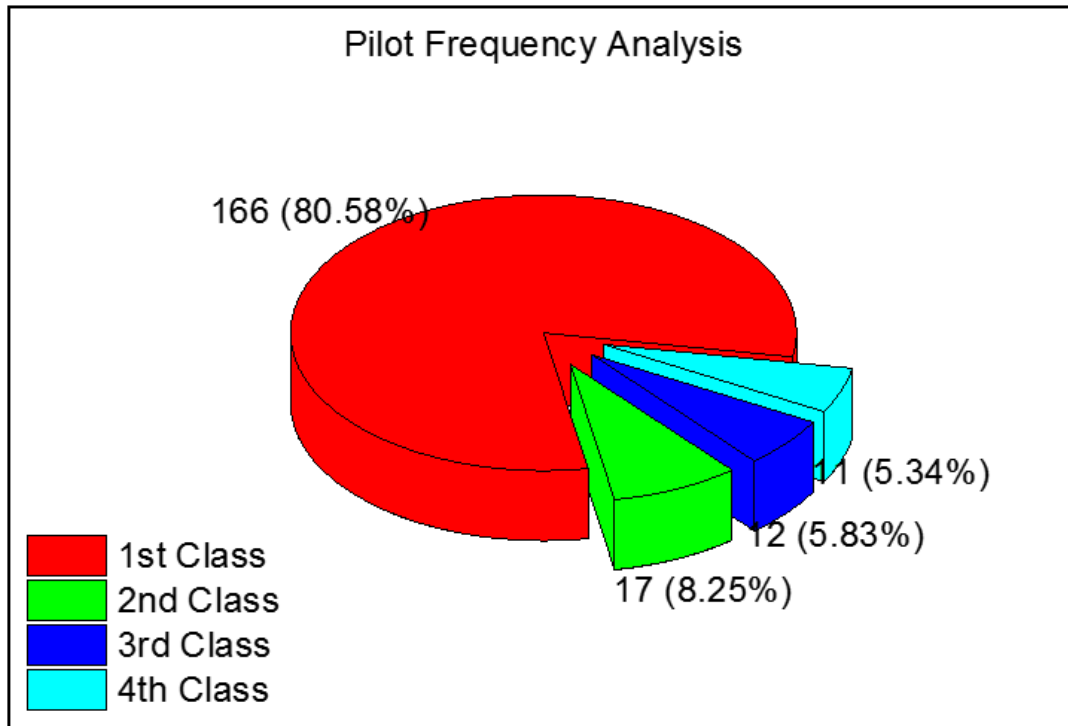


Fig. 4.8 Pilot Frequency Analysis

Table 4.7 Cross analysis of JettyNo.-Berthing Velocity

(N=206)

		JettyNo.			χ^2
		Jetty 1	Jetty 2	Jetty 3	
Berthing Velocity	Over10	13 (26.0)	26 (52.0)	11 (22.0)	14.519***
	Under10	72 (46.2)	37 (23.7)	47 (30.1)	
Total		85 (41.3)	63 (30.6)	58 (28.2)	

$p^* < .1$, $p^{**} < .05$, $p^{***} < .01$

Jetty No.를 접안속도 10cm/s 이상인 경우와 10cm/s 미만인 경우로 분리하여 교차분석을 실시하였다(Table 4.7). 그 결과 $\chi^2 = 14.519$, $p = 0.001$ 으로 접안속도에 따른 Jetty No.의 차이가 유의하게 나타났다. 접안속도 10cm/s 이상의 경우 ‘Jetty No.2’ 26회(52.0%), ‘Jetty No.1’ 13회(26.0%), ‘Jetty No.3’ 11회(22.0%) 순으로 높은 빈도를 보였다. 반면 접안속도 10cm/s 이하인 경우 ‘Jetty No.1’ 72회(46.2%), ‘Jetty No.3’ 47회(30.1%), ‘Jetty No.2’ 37회(23.7%) 순으로 많은 빈도를 보였다. 접안속도 10cm/s 미만인 경우에는 Jetty No. 전체의 빈도분석과 비슷한 경향을 보이거나 10cm/s 이상인 경우에는 Jetty No. 2의 빈도가 52%로 비교적 높은 비율을 차지하고 있다(Fig 4.9, Fig 4.10).

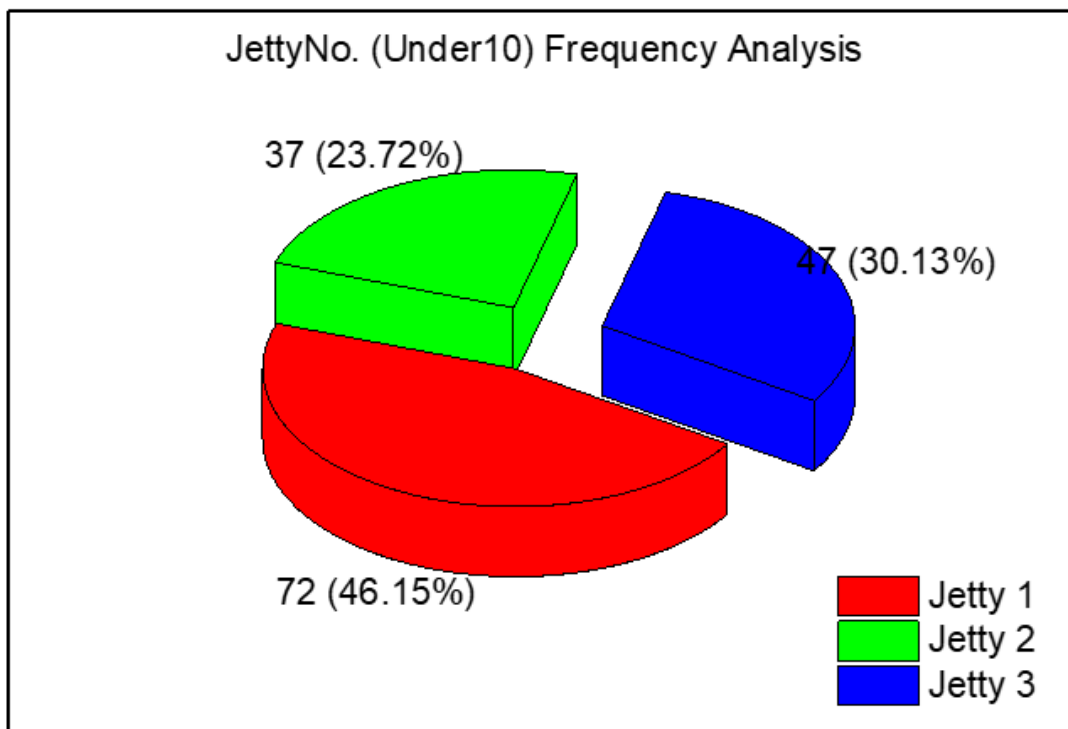


Fig. 4.9 Jetty No.(Under10) Frequency Analysis

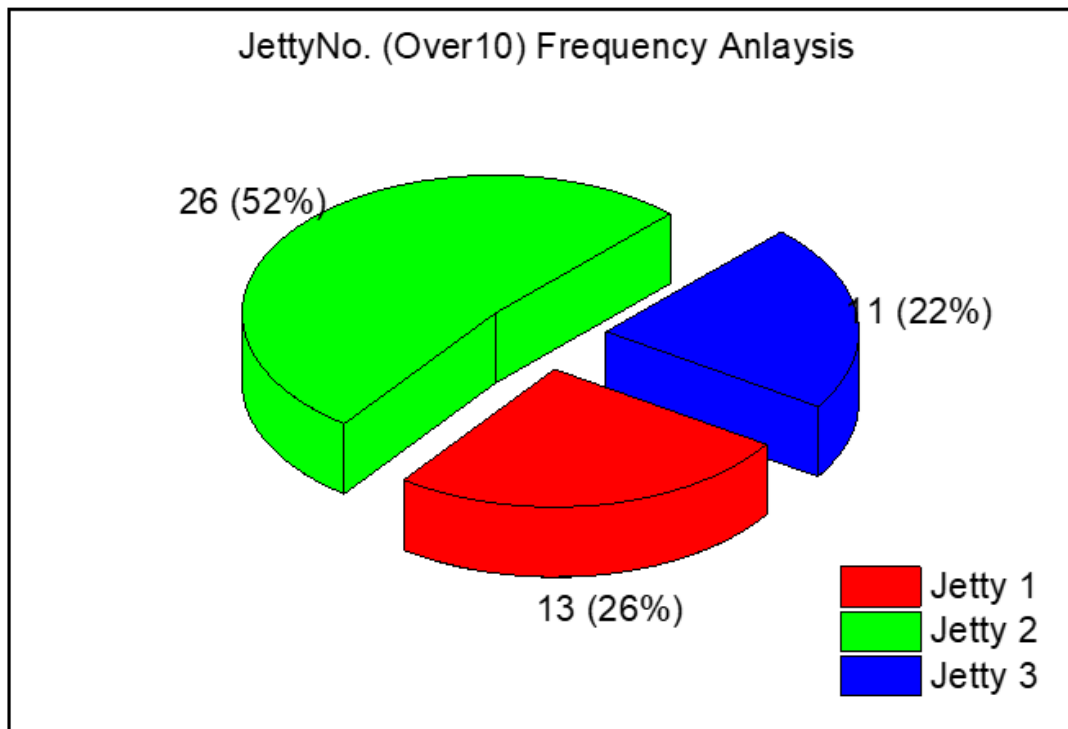


Fig. 4.10 Jetty No.(Over10) Frequency Analysis

Table 4.8 Cross anlaysia of STATE-Berthing Velocity

(N=206)

		STATE		χ^2
		Ballast	Laden	
Berthing Velocity	Over10	33 (66.0)	17 (34.0)	18.140***
	Under10	50 (32.1)	106 (67.9)	
Total		83 (40.3)	123 (59.7)	

$p^* < .1$, $p^{**} < .05$, $p^{***} < .01$

STATE를 접안속도 10cm/s를 기준으로 분리하여 교차분석을 실시하였다 (Table 4.8). 그 결과 $\chi^2 = 18.140$, $p = 0.000$ 으로 접안속도에 따른 STATE의 교차분석은 유의하게 나타났다. 접안속도 10cm/s 이상의 경우 ‘Ballast’ 33회 (66.0%), ‘Laden’ 17회(34.0%) 순으로 나타났으며, 접안속도 10cm/s 이하인 경우 ‘Laden’ 106회(67.9%), ‘Ballast’ 50회(32.1%)으로 나타나 확연한 차이를 보였다(Fig 4.11, Fig 4.12).

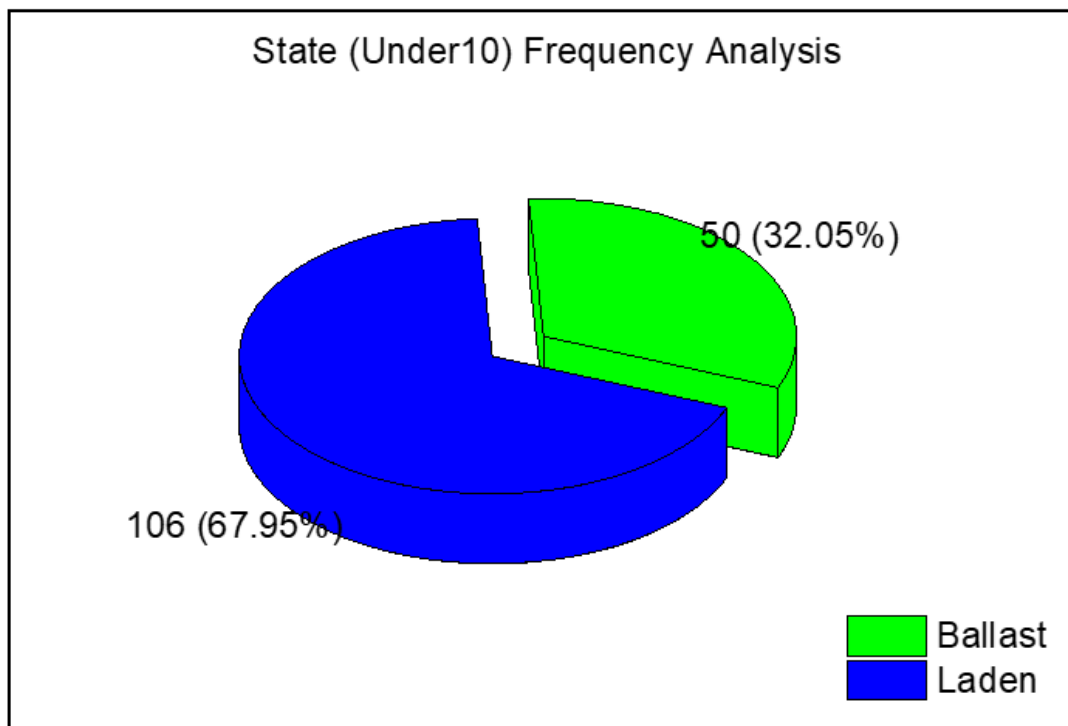


Fig. 4.11 STATE(Under10) Frequency Analysis

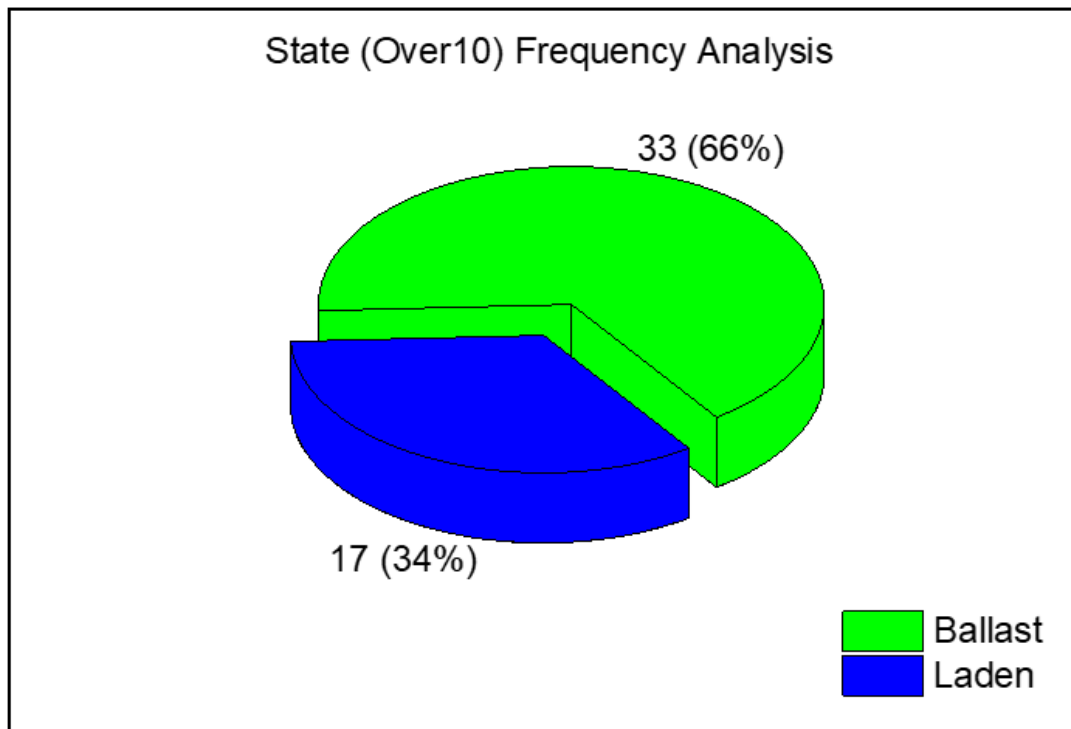


Fig. 4.12 STATE(Over10) Frequency Analysis

Table 4.9 Cross analysis of Pilot-Berthing Velocity

(N=206)

		Pilot				χ^2
		1 st Class	2 nd Class	3 rd Class	4 th Calss	
Berthing Velocity	Over10	38 (76.0)	7 (14.0)	3 (6.0)	2 (4.0)	3.040
	Under10	128 (82.1)	10 (6.4)	9 (5.8)	9 (5.8)	
Total		166 (80.6)	17 (8.3)	12 (5.8)	11 (5.3)	

$p^* < .1$, $p^{**} < .05$, $p^{***} < .01$

마지막으로 독립변수 Pilot을 접안속도 10cm/s 이상인 경우와 10cm/s 미만인 경우로 분리하여 교차분석을 실시하였다(Table 4.9). 그 결과 $x^2 = 3.040$, $p = 0.385$ 로 접안속도에 따른 Pilot 변수의 차는 유의한 결과가 나타나지 않았다. 빈도분석의 결과는 접안속도 10cm/s 이상의 경우 ‘1st Class’ 38회(76.0%), ‘2nd Class’ 14회(14.0%), ‘3rd Class’ 3회(6.0%), ‘4th Class’ 2회(4.0%) 순으로 높은 빈도를 보였다. 반면 접안속도 10cm/s 이하인 경우 ‘1st Class’ 128회(82.1%), ‘2nd Class’ 10회(6.4%), ‘3rd Class’ 9회(5.8%), ‘4th Class’ 9회(5.8%) 순으로 많은 빈도를 보였다. Pilot 변수의 경우 10cm/s 미만, 10cm/s 이상 모두 1급 도선사의 비율이 높으며 이것은 전체 Pilot의 빈도분석과 같은 경향을 보인다(Fig 4.13, Fig 4.14).

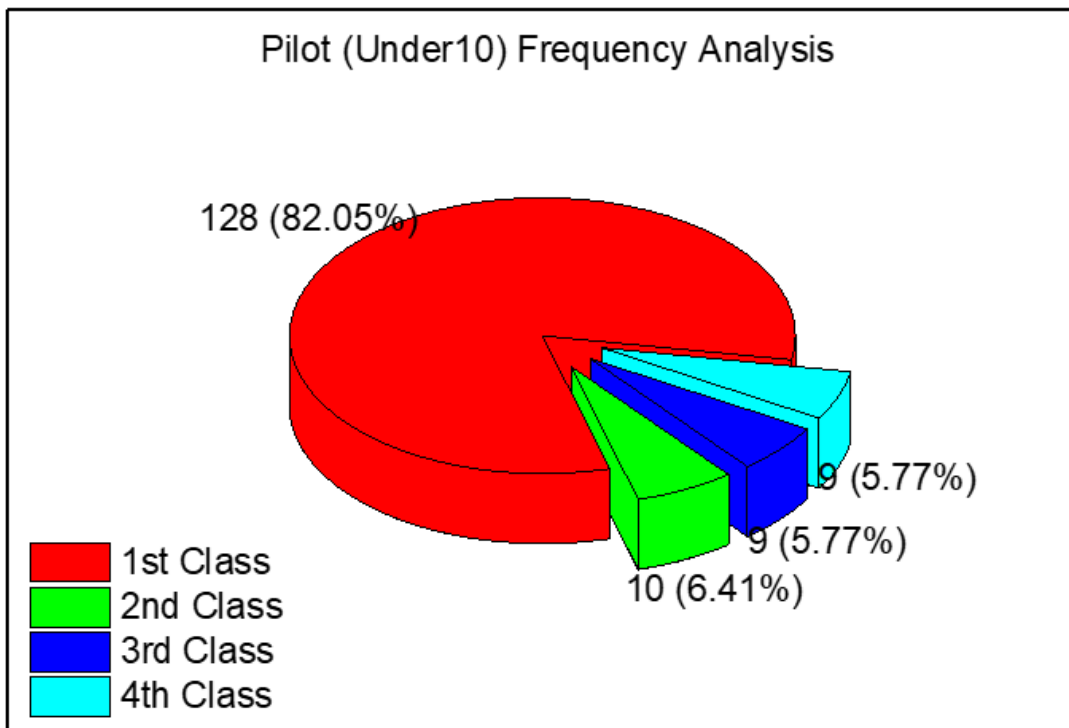


Fig. 4.13 Pilot(Under10) Frequency Analysis

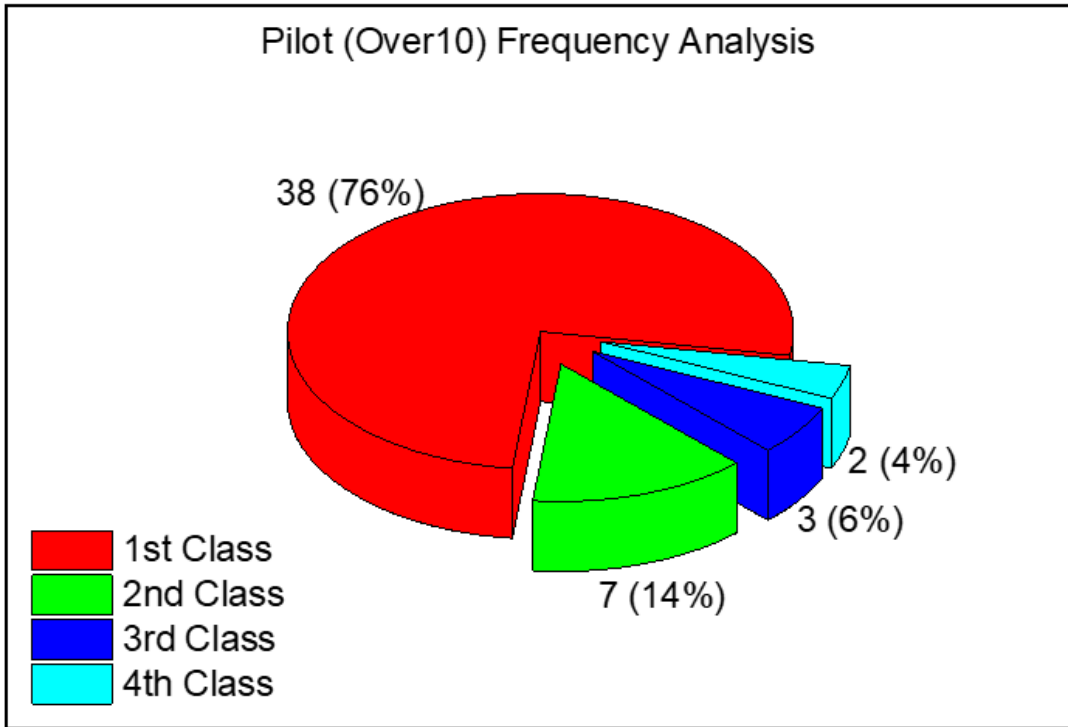


Fig. 4.14 Pilot(Over10) Frequency Analysis

4.2.2 연속형 변수

연속형 변수(Continuous variable)은 숫자형 변수이므로 Scatter Graph를 통하여 추세를 알 수 있다. 따라서 종속변수인 점안속도를 y축으로 하는 Scatter Graph를 보면 다음과 같다.

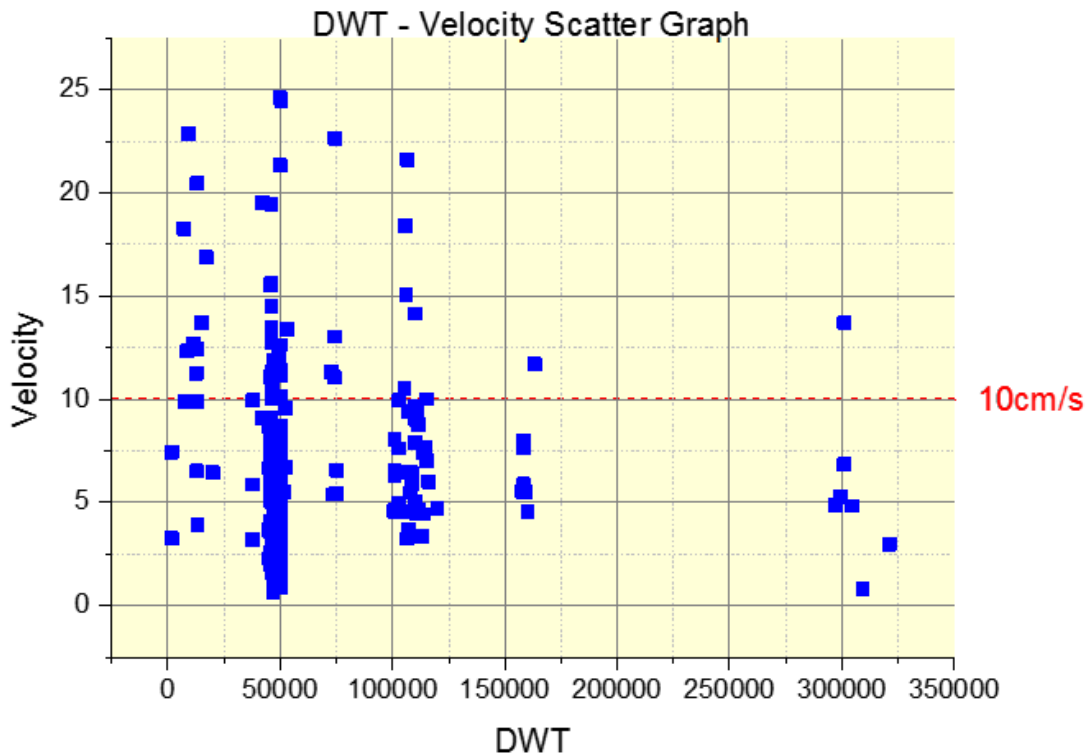


Fig. 4.15 DWT-Velocity Scatter Graph

Scatter Graph를 통한 DWT와 접안속도 간의 관계(Fig. 4.15)를 살펴보면 R^2 (R square) 값은 -0.14177로서 설명력이 매우 낮다. 그러나 음수의 추세를 보이면서 미세하나마 DWT가 클수록 접안속도 값은 낮은 것을 알 수 있다. 이는 선박의 질량이 클수록 접안에너지 값이 높아지기 때문에 대체적으로 낮은 접안속도를 유지하며 선박이 접안함을 알 수 있다.

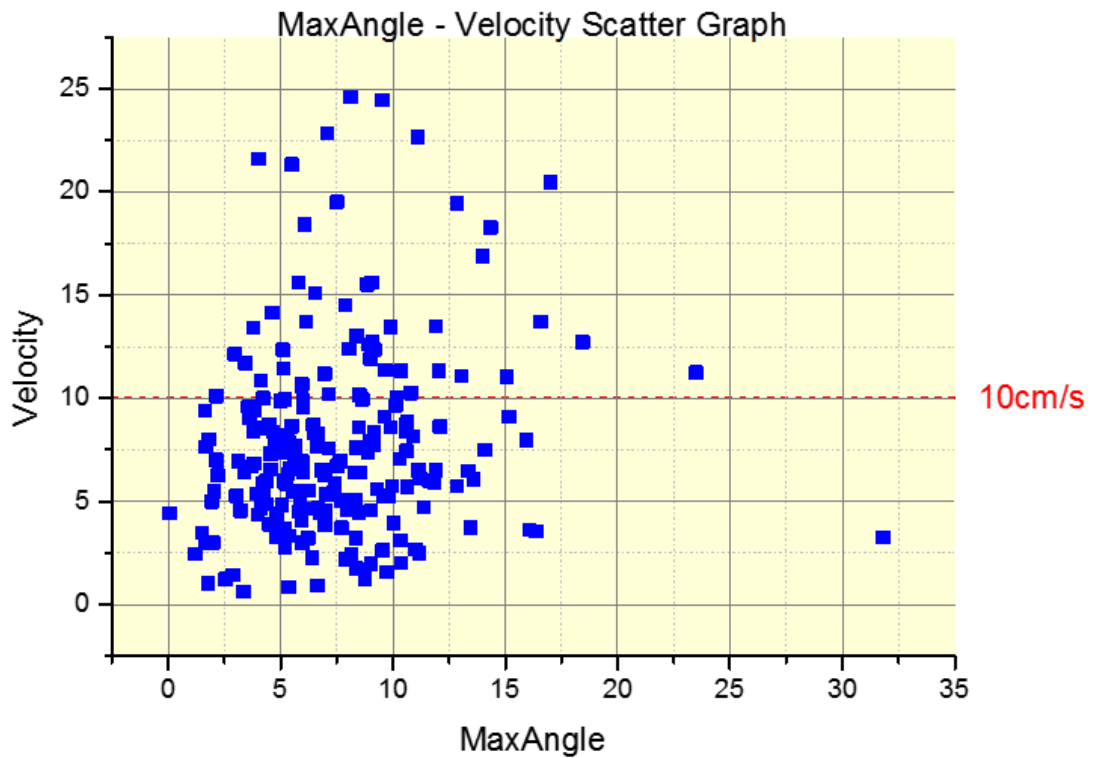


Fig. 4.16 MaxAngle-Velocity Scatter Graph

MaxAngle(접안각도)와 접안속도 간의 관계도 마찬가지로 R^2 값이 0.16414로 설명력이 매우 낮다. 하지만 미세하게 접안각도가 클수록 접안속도가 높아지는 경향이 보인다(Fig. 4.16).

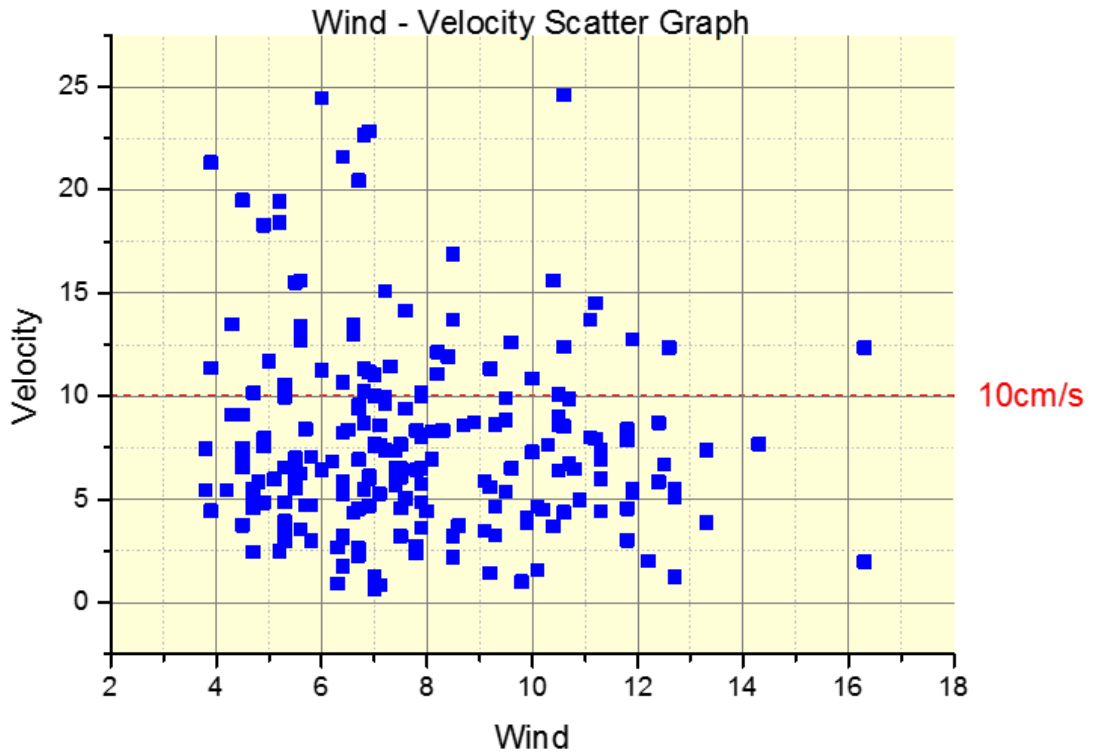


Fig. 4.17 Wind-Velocity Scatter Graph

Wind(최대풍속)와 접안속도 간의 관계는 R^2 값이 -0.1051 로 거의 관계가 없다. 그러므로 바람과 접안속도 사이의 경향은 뚜렷하게 보이지 않는다고 할 수 있다(Fig 4.17).

따라서 DWT, MaxAngle, Wind의 변수와 접안속도와의 관계에 따르면 각각의 변수는 접안속도에 큰 영향력을 미치지 않는다고 할 수 있다. 이는 선박 접안 시 도선사의 성향 등 규칙적이지 않은 요소들이 많기 때문에 이러한 경향을 나타내는 것으로 사료된다. 따라서 선박의 접안속도를 분석할 시, 고려해야할 요소가 많기 때문에 모든 변수를 포함하여 복합적으로 분석이 이루어져야할 필요성이 있다.

4.3 연관성분석

데이터 분석을 위해서는 변수의 개수 및 형태에 따라 분석기법을 다르게 적용한다. 독립변수의 수에 따라 단변량(One variable), 다변량(Multi variables)으로 나뉜다. 본 연구는 6개의 독립변수를 가지고 분석을 진행하므로 다변량 데이터 분석이다. 다변량 데이터는 변수와 변수 사이의 관계에 의해 분석이 이루어지므로 이에 따른 연관성분석을 수행해야한다. 연관성분석 후 유의미한 변수를 선택하고 무의미한 변수는 제거가 가능하다. 독립변수의 척도에 따라 연속형 변수의 경우 상관분석으로 연관성 정도를 분석하며 범주형 변수의 경우 Phi 계수와 Cramer's V 계수를 통한 상관분석 또는 교차분석으로 변수간 독립성 여부를 파악한다.

접안속도 실측데이터의 연관성분석을 위해서 IBM SPSS Statistics version 25 프로그램을 이용하였다. 변수는 연속형변수, 범주형변수로 구분하여 상관분석, 교차분석을 실시하였다. 분석 후 유의하지 않은 독립변수는 제외하여 필요한 변수를 선별하였다.

4.3.1 상관분석

상관분석(Correlation analysis)은 변수 간의 선형관계 정도를 분석하여 연관성을 파악하기 위해 사용되는 분석기법 중 하나이다. 상관분석은 두 변수 간의 공통된 분포를 나타내는 분산인 공분산(covariance)을 이용한다. 이 때, 측정단위에 따라 변하지 않는 두 변수 간의 선형관계를 나타내는 표준화된 공분산을 도출할 수 있는데, 이 값을 상관계수라고 한다. 이와 같이 두 변수의 표준화된 편차곱의 평균으로 구한 상관계수를 피어슨의 적률상관계수(Pearson's product-moment correlation coefficient) 또는 피어슨 상관계수라 한다(이훈영, 2008).

상관계수는 두 변수간의 선형관계 정도와 방향을 수학적 수치로 정량화하여 표시하는 지수로 -1에서 +1 사이의 값을 가진다. 1에 가까울수록 양적상관관계, -1에 가까울수록 음적상관관계를 가진다.

상관계수는 부호에 관계없이 상관계수의 절대값 크기가 변수들 간의 연관성 정도를 판단하는 기준이 되는데 어떠한 자료의 변수 간 연관성분석을 실시하는가에 따라 상관 정도를 판단하는 기준은 서로 다를 수 있다. 하지만 일반적인 상관계수의 해석은 Table 4.10와 같다.

Table 4.10 Qualitative interpretation of Correlation coefficient

Absolute Value of the Correlation Coefficient	Strength of Linear Relationship
0.0 ~ 0.2	Very weak
0.2 ~ 0.4	Weak
0.4 ~ 0.6	Moderate
0.6 ~ 0.8	Strong
0.8 ~ 1.0	Very Strong

또한 상관분석을 통하여 도출한 상관계수가 통계적으로 유의한가를 검정할 필요가 있다. 통계적으로 유의하여야 변수 간의 상관계수를 일반화하여 사용할 수 있다. 따라서 상관계수와 유의성을 모두 고려한 상관분석의 결과는 Table 4.11과 같다.

Table 4.11 Correlation Analysis of Continuous Variables

(N=206)

	Velocity	DWT	MaxAngle	Wind
Velocity	1	-.142*	.164*	-.105
DWT		1	-.390**	-.106
MaxAngle			1	-.182**
Wind				1

$p^* < .1$, $p^{**} < .05$, $p^{***} < .01$

접안속도와 각 변수 별 상관분석 결과 Velocity와 DWT 값은 $r=-0.142$ 이며 $p=0.042$ 로서 음의 상관관계를 가지며, MaxAngle 값은 $r=0.164$, $p=0.028$ 로서 양의 상관관계를 가진다. 상관계수 값이 크지 않지만 약한 상관관계가 있다고 분석되었다. 하지만 Wind의 경우 $r=-0.105$ 이나 $p=0.133$ 으로 유의하지 않는 것으로 나타났다. DWT와 MaxAngle의 상관관계는 $r=-0.390$, $p=0.000$ 으로 음의 상관관계를 가지며, DWT와 Wind의 상관관계는 $r=-0.106$, $p=0.128$ 로서 유의하지 않다.

범주형변수의 상관관계 분석은 변수의 개수에 따라 Phi 계수와 Cramer's V 계수를 사용하여 분석한다. Phi 계수는 비교대상 변수가 남, 여와 같이 2개인 경우 사용한다. 본 연구에서는 Velocity와 STATE의 상관관계를 분석할 때 두 변수 모두 2개의 비교대상 범주를 가지므로 Phi 계수를 적용하였다. Cramer's V 계수의 경우 비교대상 범주가 3개 이상인 경우 사용되는데, Velocity와 Jetty No., Pilot의 경우에 사용하였다.

Table 4.12 Correlation Analysis of Categorical Variables

Velocity - Jetty No.			Value	Aprrox. Sig
	Categorical by Categorical	Phi	0.267	0.001
		Cramer's V	0.267	0.001
	N of Vaild Cases			212
Velocity - STATE			Value	Aprrox. Sig
	Categorical by Categorical	Phi	-0.274	0.000
		Cramer's V	0.274	0.000
	N of Vaild Cases			212
Velocity - Pilot			Value	Aprrox. Sig
	Categorical by Categorical	Phi	0.110	0.465
		Cramer's V	0.110	0.465
	N of Vaild Cases			212

Table 4.14에 따르면 Velocity와 Jetty No.의 경우 Cramer's V 계수 값이 0.267이며 $p=0.001$ 로서 낮은 상관관계를 가진다. STATE의 경우 Phi 계수 값은 -0.274, $p=0.000$ 으로 역시 Velocity와의 상관관계를 가진다. Pilot의 경우 $p=0.465$ 로서 유의하지 않는 것으로 나타났다.

4.3.2 교차분석

교차분석은 명목척도나 서열척도로 측정된 범주형 변수들 간의 범주를 교차시켜 얻은 빈도교차표를 이용하여 두 변수 간의 상호독립성 및 관련성 정도를 분석하는 방법이다(이훈영, 2008).

우선 Jetty No.와 STATE 간의 교차분석을 통해 연관성을 분석하였다. 분석 결과 $\chi^2 = 24.290$, $p = 0.000$ 으로 Jetty No.와 STATE 변수 사이는 독립적이지 않음을 알 수 있다(Table 4.13).

Table 4.13 Cross analysis of JettyNo.-STATE

(N=206)

		JettyNo			χ^2
		Jetty 1	Jetty 2	Jetty 3	
STATE	Ballast	28 (33.7)	41 (49.4)	14 (16.9)	24.290***
	Laden	57 (46.3)	22 (17.9)	44 (35.8)	
Total		85 (41.3)	63 (30.6)	58 (28.2)	

$p^* < .1$, $p^{**} < .05$, $p^{***} < .01$

다음은 Jetty No.와 Pilot 간의 교차분석을 통해 연관성을 분석하였다. 분석 결과 $\chi^2 = 6.177$, $p = 0.404$ 으로 Jetty No.와 Pilot 변수 사이는 독립적이라는

귀무가설을 채택하게 된다(Table 4.14).

Table 4.14 Cross analysis of JettyNo.-Pilot

(N=206)

		JettyNo			x^2
		Jetty 1	Jetty 2	Jetty 3	
Pilot	1 st Class	69 (41.6)	48 (28.9)	49 (29.5)	6.177
	2 nd Class	9 (52.9)	7 (41.2)	1 (5.9)	
	3 rd Class	3 (25.0)	5 (41.7)	4 (33.3)	
	4 th Class	4 (36.4)	3 (27.3)	4 (36.4)	
Total		85 (41.3)	63 (30.6)	58 (28.2)	

$p^* < .1$, $p^{**} < .05$, $p^{***} < .01$

마지막으로 STATE와 Pilot 간의 교차분석을 통해 연관성을 분석하였다. 분석 결과 $x^2 = 2.148$, $p = 0.542$ 으로 STATE와 Pilot 변수 사이는 앞서 Jetty No.와 Pilot 변수 사이와 마찬가지로 변수 사이는 독립적이라는 귀무가설을 채택하게 된다(Table 4.15).

Table 4.15 Cross analysis of STATE-Pilot

(N=206)

		STATE		χ^2
		Ballast	Laden	
Pilot	1 st Class	67 (40.4)	99 (59.6)	2.148
	2 nd Class	9 (52.9)	8 (47.1)	
	3 rd Class	4 (33.3)	8 (66.7)	
	4 th Class	3 (27.3)	8 (72.7)	
Total		83 (40.3)	123 (59.7)	

$p^* < .1$, $p^{**} < .05$, $p^{***} < .01$

4.3.3 변수채택

연관성분석인 상관분석 및 교차분석 결과에 따라 본 연구에서는 6가지 종속 변수 중 유의한 관계를 가지고 있지 않은 Wind와 Pilot 변수를 제외하고 분석에 필요한 변수를 채택하였다. 따라서 분석에 사용되는 독립변수는 Jetty No., STATE, DWT, MaxAnlge이다(Table 4.16).

Table 4.16 Selection Variables

(N=206)

Independent Variable	Type	Description	Mean
Jetty No.	Categorical (Nominal scale)	1 = Jetty 1 2 = Jetty 2 3 = Jetty 3	1.879
STATE	Categorical (Nominal scale)	1, if Ballast; 0, otherwise	0.597
DWT	Continuous (Interval scale)	Deadweight tonnages	70088.762
MaxAngle	Continuous (Interval scale)	Ship's berthing Angle (degrees)	7.519

Dependent Variable	Type	Description	Mean
Berthing Velocity	Continuous (Interval scale)	Max ship's Berthing Velocity (cm/s)	7.858
	Categorical (Nominal scale)	1, if Under10; 0, otherwise	0.242

4.4 다중선형회귀

회귀(regression)분석은 종속변수와 독립변수 사이의 관계를 분석하여 결과를 나타내는 것이다. 회귀분석은 독립 값과 의존 값 사이의 관계를 직선으로 가정해 시작한다.

y는 종속변수, x는 독립변수인 $y = ax + b$ 같은 기울기 절편으로 정의한 직선이 있을 때 기울기 a는 x에 대해 직선이 얼마나 증가하는가를 나타내며 b는 x값이 0일 때 y값을 나타낸다. 회귀는 기울기 절편과 유사한 형태를 사용하여

데이터를 모델화하는 방식이며, 명시된 직선으로 x 값을 y 로 가장 잘 표현하는 a 와 b 의 값을 구하는 방식이다. 회귀분석은 다시 단순선형회귀 분석과 다중선형회귀 분석으로 구분된다(Brett Lantz, 2014). 다중선형회귀 분석은 다수의 종속변수의 상관관계를 통하여 통계적 방법에 의해 추정하는 분석방법이다.

다중선형회귀의 목표는 실제 y 값과 예측된 y 값 사이의 오차를 제곱한 합을 최소로 만드는 벡터 β 를 구하는 것이다. 최적의 해답을 위해서 선형대수가 필요하며 벡터 β 의 최적 추정은 다음과 같다(박승수, 2016).

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (4.1)$$

식 3.1에서 구한 β 계수의 값을 가지고 다음 방정식(4.2)으로 다중회귀방정식을 구할 수 있다. 독립변수 y 는 절편 항의 총합과 i 번째 속성에 대한 x 값과 추정된 β 값의 곱을 다 더하고 추가하여 오차항(ε)을 더한 것으로 나타낸다.

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i + \varepsilon \quad (4.2)$$

x_i 의 변화에 따라 β_i 만큼 y 값은 변화한다. 따라서 각 속성은 y 값에 대한 추정된 효과를 가진다고 할 수 있다. 또한 독립변수가 모두 0인 경우의 절편은 y 의 예상값이므로 α 는 다른 회귀변수와 다르지 않기 때문에 식 4.3과 같이 나타낼 수 있다.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i + \varepsilon \quad (4.3)$$

따라서 다중선형회귀 공식은 식 4.4와 같이 간결하게 나타낼 수 있다.

$$Y = X\beta + \varepsilon \quad (4.4)$$

김호영(2005)은 자동차의 주행속도 예측모형 구축을 위해 국내의 한 고속도로를 대상으로 주행속도에 미치는 영향을 고려하여 데이터를 수집하였다. 고속도로 구간별 수집된 데이터를 통해 다중선형회귀 모형을 구축하여 주행속도 예측식을 제안하였다. 이와 같이 본 연구에서도 선박 접안 실측데이터를 이용하여 다중선형회귀분석을 통하여 선박 접안속도 예측식을 도출하고자 한다.

다중선형회귀분석에는 IBM SPSS Statistics version 25 프로그램을 이용하였다. 상관관계 분석을 바탕으로 선별한 Jetty No., DWT, STATE, MaxAngle를 종속변수, Velocity를 독립변수로 하여 분석을 실시하였다. 또한 다중선형회귀분석을 위하여 범주형변수를 더미변수(dummy variable)로 변환하였다. 더미변수는 변수를 0과 1로 인위적으로 코딩한 변수를 의미한다. 따라서 회귀분석에 적합하지 않은 범주형변수를 더미변수로 변환하는 것은 분석을 최적화를 하는데 적합하다고 할 수 있다. 범주형변수의 더미변수 변환 결과는 Table 4.17과 같다.

Table 4.17 Dummy Variable Transformation

	Dummy1	Dummy2	Dummy3
Jetty No.1	1	0	0
Jetty No.2	0	1	0
Jetty No.3	0	0	0
STATE Laden	1	0	
STATE Ballast	0	0	

Table 4.18 Result of Multiple Regression

Model	R	R Square	Adjust R Square	F	Sig.
		0.374	0.140	0.118	6.507
Multiple Regression Analysis	Unstandradized Coefficient		Standradized Coefficient	t	Sig.
	B	Std. Error	Beta		
(Constant)	10.933	1.475		7.411	0.000
Jetty No.1	-2.786	1.025	-0.296	-2.717	0.007
Jetty No.2	-0.986	1.141	-0.098	-0.864	0.388
DWT	-1.642E-05	0.000	-0.193	-1.936	0.054
State Laden	-2.064	0.662	-0.219	-3.116	0.002
MaxAngle	0.100	0.084	0.087	1.187	0.237

다중선형회귀분석 결과(Table 4.18) 따르면 모형의 R^2 의 값이 0.140으로 매우 낮아 통계적으로 유의하지 않다고 할 수 있다. 또한 유의하지 않은 변수가 존재하기 때문에 유용한 독립변수 선택을 위해 전진선택방식(forward selection)과 후진제거방식(backward elimination)을 결합한 단계선택방식(stepwise selection)을 통하여 새로운 다중선형회귀분석을 실시하였다.

Table 4.19 Result of Multiple Regression after stepwise selection

Model	R	R Square	Adjust R Square	F	Sig.
		0.363	0.132	0.119	10.233
Multiple Regression Analysis	Unstandradized Coefficient		Standradized Coefficient	t	Sig.
	B	Std. Error	Beta		
(Constant)	10.997	0.672		16.360	0.000
Jetty No.1	-2.248	0.671	-0.239	-3.349	0.001
State Laden	-1.968	0.651	-0.208	-3.022	0.003
DWT	-1.461E-05	0.000	-0.175	-2.410	0.017

단계선택방식을 통해 유의미한 변수를 선택하여 실시한 다중선형회귀분석 결과(Table 4.19)에 따르면 모형의 R^2 의 값이 0.132으로 역시 매우 낮아 통계적으로 유의하지 않은 결과가 나왔다. 이는 분석의 결과를 이용하여 선박 접안속도 예측식을 도출하여도 정확도가 매우 낮다는 것을 의미한다. 따라서 다중선형회귀 분석결과로는 접안속도 예측이 어려우며, 접안속도에 영향을 미치는 요인 파악이 어려우므로 본 연구에서는 머신러닝 기법을 활용하여 접안속도 예측모형을 구축하고자 한다.

제 5 장 예측모델 알고리즘

5.1 개요

예측모델 알고리즘을 구축하기 위해 데이터 분석에 사용된 프로그래밍 언어, IDE(Integrated development environment)는 Table 5.1과 같다. 프로그래밍 언어는 R을 사용하였으며, 통합개발환경(IDE)은 R Studio Server를 설치하여 원격에서 웹으로 서버에 접속하는 분석방법을 사용하였다.

Table 5.1 Analysis environment

	Name	Version
A programing language	R	3.5.1
IDE	R sstudio Server	1.1.456

R은 무료 오픈소스 소프트웨어로서 GPL 라이선스를 적용받았다. 전 세계의 연구자들이 통계, 데이터마이닝, 머신러닝, 시각화 도구 등의 알고리즘을 개발하고 공유한다. R에서의 확장 기능은 Package라 하며, 본 연구에서 사용된 R package는 Table 5.2와 같다.

Table 5.2 Machine Learning algorithm package in R

Algorithm	Package	Version
C5.0	C50	0.1.2
RandomForest	randomForest	4.6-14
Logistic Regression	glmnet	2.0-16
Artificial Neural Network	nnet	7.3-12

R package에서 각 알고리즘을 분석할 시 목적함수의 결과를 최대로 만드는 매개변수를 설정하는데, 이 때 사용자가 직접 설정해야하는 매개변수를 머신러닝에서는 하이퍼파라미터(Hyperparameter)라고 한다. 설정된 알고리즘 별 하이퍼파라미터는 Table 5.3과 같다. C5.0은 예측 정확도를 높이기 위해 부스팅 기능이 내장되어 있지만 사용하지 않았다. 랜덤포레스트의 경우 50개의 모델을 조합하였다. 로지스틱 회귀분석을 위해 하이퍼파라미터는 binomial로 설정하였다. 인공신경망의 경우 size = 10, maxit = 1000를 입력하여 학습 효율을 높였다.

Table 5.3 Hyperparameter

Algorithm	Hyperparameter
C5.0	trials = 1
RandomForest	ntree = 50 mtry = 2 importance = T
Logistic Regression	family = binomial
Artificial Neural Network	size = 10 maxit = 1000 decay = 0.1 rang = 0.1 bag = False

예측모델 알고리즘의 예측정확도 검증을 위한 방식으로는 Hold-Out 방식을 채택하였다. Hold-Out 방식은 데이터를 Train set와 Test set로 구분하며 Train set에서 모델을 학습시키고 Test set을 이용하여 예측한다. Test set 정확히 예측된 비율이 예측 모델의 정확도에 대한 추정치라고 할 수 있다. Hold-Out 검증방법을 이용한 전체과정을 그림으로 요약하면 Fig 5.1과 같다.

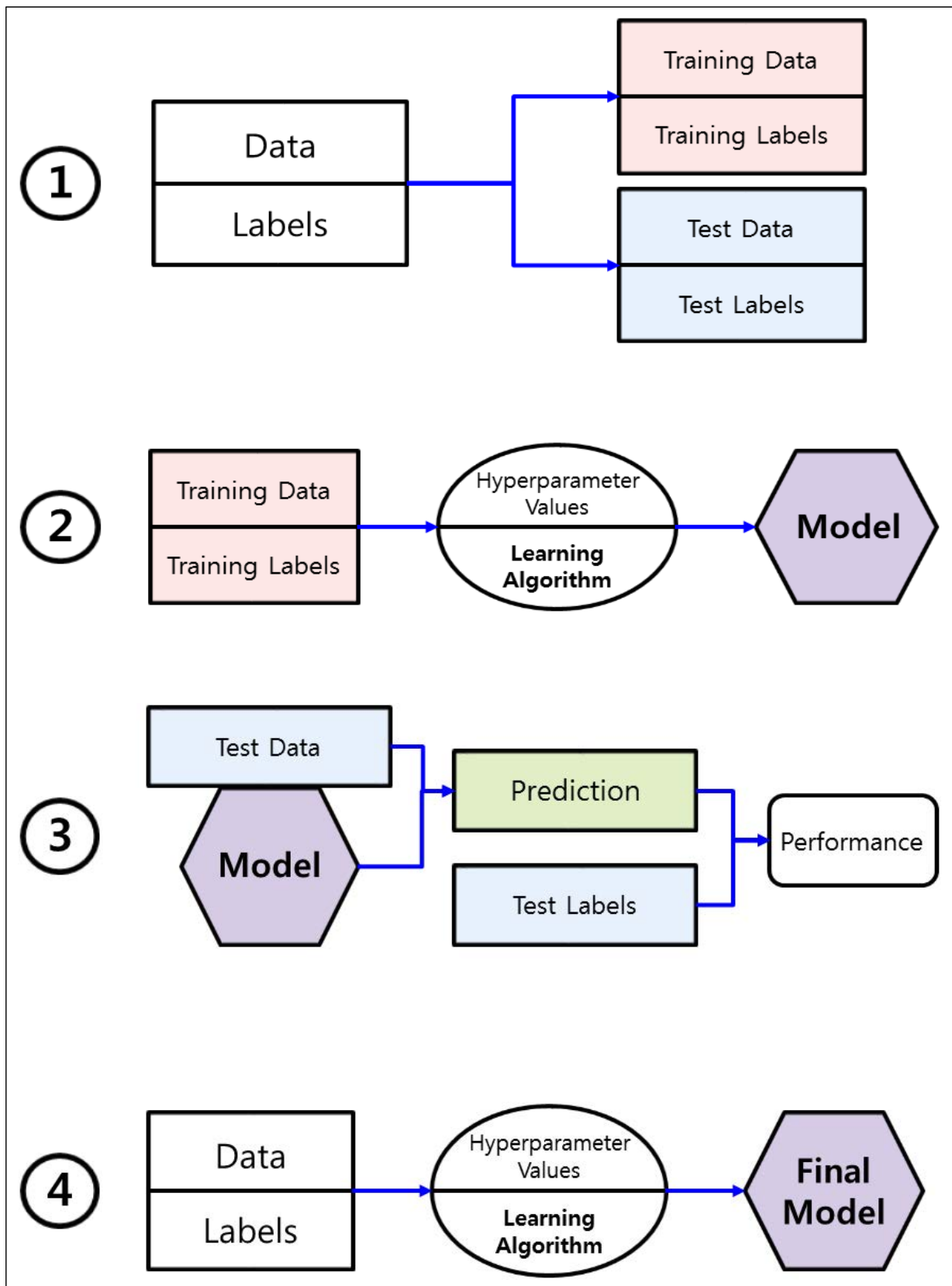


Fig. 5.1 Hold-Out method

첫 번째 단계에서는 전체데이터를 Train set와 Test set으로 랜덤하게 나눈다. 여기서 Test set는 학습 알고리즘을 위한 새로운 데이터의 역할을 하게 된다. 보통 2/3을 Train set로 1/3을 Test set로 설정한다.

두 번째 단계는 적합한 학습 알고리즘을 선택하는 일이다. 여기서 하이퍼파라미터를 지정하고 모델을 학습시킨다.

세 번째 단계는 Test 단계로서, Test set을 통하여 모델에 대한 성능을 추정한다. 여기서 학습된 알고리즘을 통하여 Test set의 값을 예측하게 된다.

마지막으로 전체 데이터를 통하여 다시 학습 알고리즘을 적용시켜 최종 모델을 완성한다.

따라서 본 연구에서는 206개의 전체데이터를 7:3의 비율로 Train data set(144개)과 Test data set(62개)을 구분하였다. 각 알고리즘으로는 의사결정나무, 랜덤 포레스트, 로지스틱 회귀분석, 인공신경망을 사용하였으며 Training data set을 이용해 각 알고리즘별 모델 구축 후 Test data set을 이용하여 성능비교 및 검증 하였다.

5.2 접안속도 예측모델

접안속도 예측모델 구축을 위하여 수집된 206개의 실측데이터를 바탕으로 구분한 Train data set(144개)을 사용하였다.

5.2.1 의사결정나무

Fig. 5.2은 의사결정나무 C5.0 알고리즘을 적용하여 R을 이용해 예측모델을 구축한 것이다. 의사결정나무 알고리즘으로 예측모델 구축결과 10개의 노드로 구성된다. 종속변수 STATE, Jetty No. DWT, MaxAngle 순서로 선정되어 추론되는 구조를 보여주고 있다. 따라서 의사결정나무 알고리즘에 의한 각 종속변수별 중요도는 노드의 순서에 따라 STATE -> Jetty No. -> DWT, MaxAngle과 같다.

의사결정나무 알고리즘 분석에 따르면 우선 STATE 변수에 의해 Laden 상태와 Ballast 상태의 경우로 분리된다. Ballast 상태인 경우 Jetty No. 변수에 따라 분리되는데 1번 Jetty의 경우 DWT 50275 ton을 기준으로 다시 분리되고 3번 Jetty의 경우 MaxAngle 9.1의 값으로 분리된다.

의사결정나무 알고리즘 적용결과 전체 학습데이터 139개 중 25개의 오류가 발생하여 17.99%의 오류율(error rate)을 보였다. Table 5.4와 같이 전체 10cm/s 이상 데이터는 35개이나 15개의 오분류를 하였고 10cm/s 이하 데이터는 104개의 데이터 중 10개를 오분류 하였다(Table 5.4).

Table 5.4 Classified Train data in Decision Tree

Classified Train	Over 10	Under 10
Over 10	20	15
Under 10	10	94

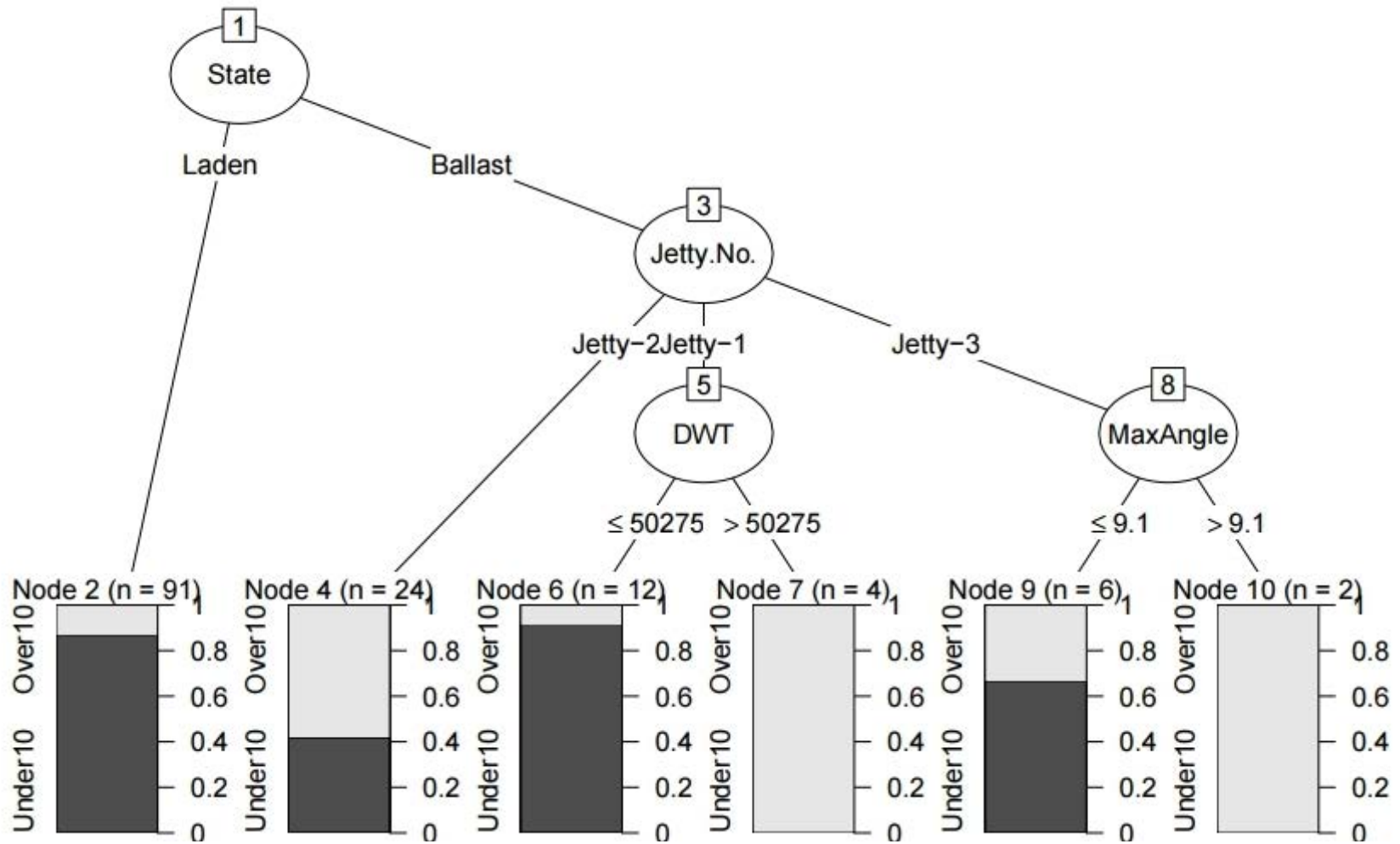


Fig. 5.2 Result of Decision Tree

5.2.2 랜덤포레스트

랜덤포레스트의 경우 알고리즘 적용결과 Fig. 5.3과 같이 50개의 나무를 구성하여 최종 모델을 완성하였을 때 오류율이 가장 낮은 것을 알 수 있었다. 오류율은 26.62%로서 분류가 비교적 완벽하게 되지는 않았다. Table 4.6과 같이 전체 10cm/s 이상 데이터는 22개의 오분류를 하였고 10cm/s 이하 데이터는 15개를 오분류 하였다(Table 5.5).

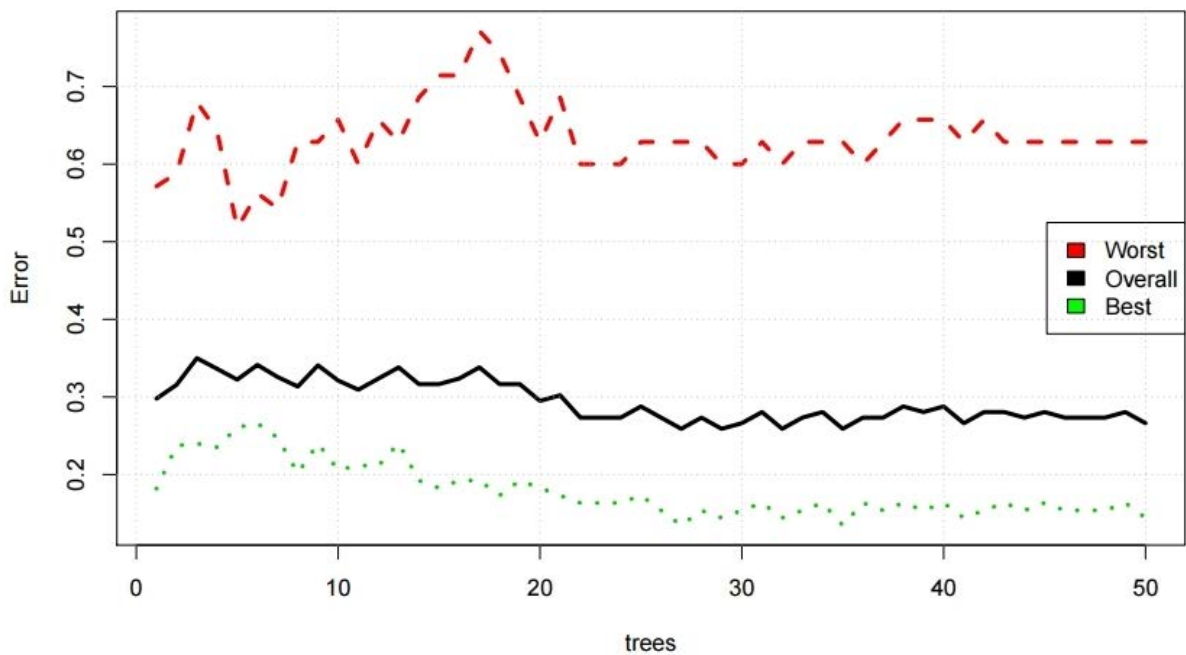


Fig. 5.3 Result of Random forest

Table 5.5 Classified Train data in Random Forest

Train \ Classified	Over 10	Under 10
	Over 10	13
Under 10	15	89

랜덤포레스트 알고리즘의 종속변수별 중요도는 Fig. 5.4과 같다. 정확도 감소성(Mean Decrease Accuracy)에 따라 종속변수의 중요도 순서를 나타내었다. 정확도 감소성은 해당 인자를 모델에서 제거하였을 때 모델의 예측정확도가 하락함에 따라 중요도를 측정한 값이다. 정확도 감소성에 따르면 종속변수의 중요도는 STATE → DWT → Jetty. No → MaxAngle 순서라고 할 수 있다.

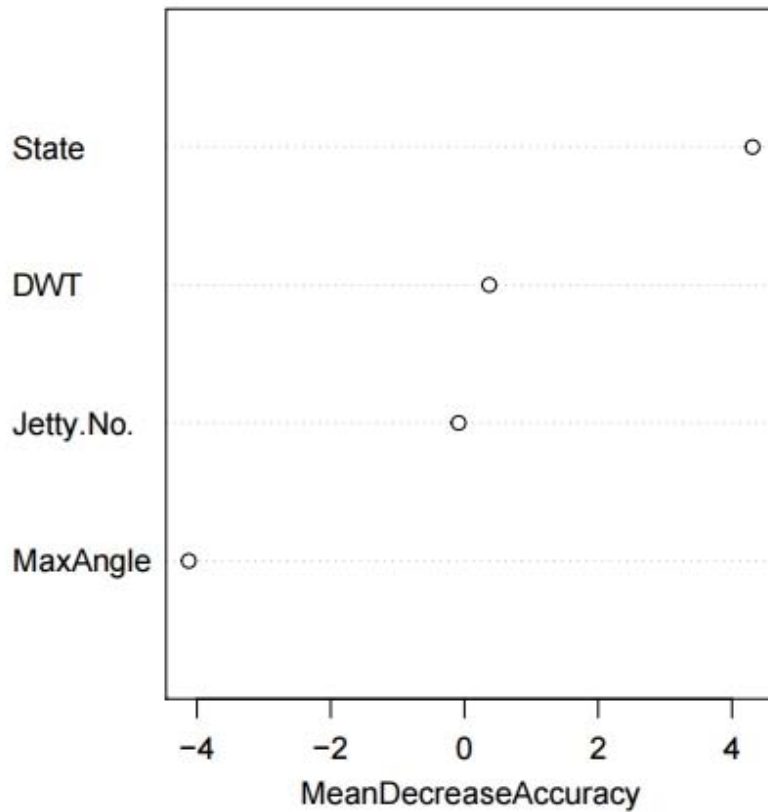


Fig. 5.4 Importance of Dependent variables in Random forest

5.2.3 로지스틱회귀

본 연구에서는 앞서 다중선형회귀분석을 실시결과 R^2 값이 통계적으로 유의하지 않을 정도로 낮아 종속변수를 범주형 이분변수로 구분(10cm/s 이상, 10cm/s 미만)하여 로지스틱 회귀분석으로 접안속도 예측모델을 구축하였다.

로지스틱 회귀분석의 결과는 Table 5.6과 같다. 분석 결과, 종속변수 중 STATE의 경우 유의수준에서 1%이하로 유의하였으나, Jetty No. DWT, MaxAngle의 경우 한계적으로 유의함을 확인하였다. 특히 변수 DWT와 MaxAngle은 접안속도에 미치는 영향이 상대적으로 낮게 나타났다.

Table 5.6 Result of Logistic Regression

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	9.151e-01	6.893e-01	1.328	0.184313
Jetty. No. 1	-	-	-	-
Jetty. No. 2	-9.692e-01	5.361e-01	-1.808	0.070620
Jetty. No. 3	-9.819e-01	7.495e-01	-1.310	0.190211
DWT	4.168e-06	6.240e-06	0.668	0.504123
STATE Ballast	-	-	-	-
STATE Laden	1.607e+00	4.583e-01	3.507	0.000453
MaxAngle	-5.204e-02	5.403e-02	-0.963	0.335414

로지스틱회귀 알고리즘 적용결과 전체 학습데이터 중 30개의 오류가 발생하여 21.58%의 오류율을 보였다. Table 5.7와 같이 전체 10cm/s 이상 데이터는 35개이나 11개의 오분류를 하였고 10cm/s 이하 데이터는 104개의 데이터 중 19개를 오분류 하였다.

Table 5.7 Classified Train data in Logistic Regression

Train \ Classified	Over 10	Under 10
	Over 10	16
Under 10	19	93

5.2.4 인공신경망

인공신경망 알고리즘 적용 결과 10개의 은닉층 노드로 구성되어 접안속도가 출력되었다(Fig. 5.5). 10개의 은닉층에는 10개의 로지스틱 함수가 있으며 출력층은 은닉층의 로지스틱 함수들을 가중결합하여 구성된다. B1, B2는 은닉층, 출력층 함수의 상수항을 의미한다. 각각의 입력층, 은닉층, 출력층간의 가중치 (weight)는 Table 5.8과 같다.

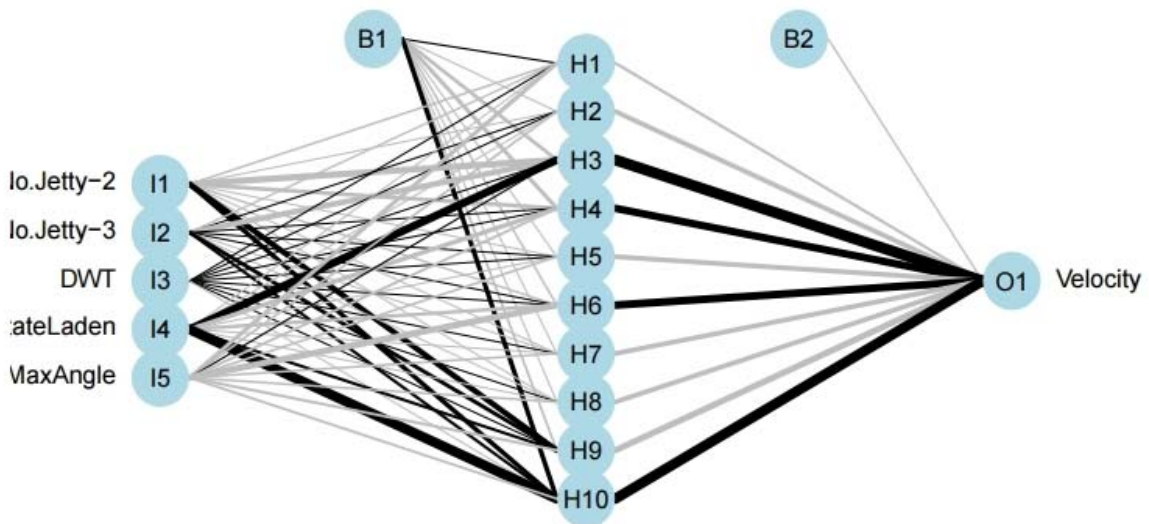


Fig. 5.5 Result of Artificial Neural Network

Table 5.8 Weights of Artificial Neural Network

	Weights					
	Input Layer - Hidden Layer weight					Hidden Layer - Output Layer weight
	I1	I2	I3	I4	I5	O1
H1	-0.15	-0.10	0.00	0.00	-0.69	-0.35
H2	-0.01	0.00	0.00	-0.01	-0.20	-0.89
H3	-1.29	-1.08	0.00	1.69	0.03	2.49
H4	-0.64	0.00	0.00	-0.78	-0.29	1.59
H5	-0.01	0.00	0.00	-0.01	-0.20	-0.89
H6	-0.13	0.00	0.00	-0.53	-1.30	1.64
H7	-0.01	0.00	0.00	-0.01	-0.20	-0.89
H8	-0.01	0.00	0.00	-0.01	-0.20	-0.89
H9	1.08	0.63	0.00	0.38	-0.34	-1.26
H10	0.80	0.38	0.00	2.21	-0.26	2.31

Table 5.9 Constant of Artificial Neural Network

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10
B1	0.04	-0.01	-0.53	-0.64	-0.01	-0.13	-0.01	-0.01	-0.23	0.83
	O1									
B2	-0.09									

인공신경망은 Table 5.9와 같이 상수항의 값과 가중치에 의해 은닉층은 로지스틱 함수를 가지게 되며 이를 결합하여 출력층에서 비선형 판별함수를 구성한다.

Table 5.10 Discriminant Function of Artificial Neural Network

		Discriminant Function
Output Layer	O1	$-0.39-(0.35 \times H1)-(0.89 \times H2)+(2.49 \times H3)+(1.59 \times H4)-(0.89 \times H5)$ $+(1.64 \times H6)-(0.89 \times H7)-(0.89 \times H8)-(1.26 \times H9)+(2.31 \times H10)$
Hidden Layer	H1	$0.04-(0.15 \times I1)-(0.10 \times I2)-(0.69 \times I5)$
	H2	$-0.01-(0.01 \times I1)-(0.01 \times I4)-(0.20 \times I5)$
	H3	$-0.53-(1.29 \times I1)-(1.08 \times I2)+(1.69 \times I4)+(0.03 \times I5)$
	H4	$-0.64-(0.64 \times I1)-(0.78 \times I4)-(0.29 \times I5)$
	H5	$-0.01-(0.01 \times I1)-(0.01 \times I4)-(0.20 \times I5)$
	H6	$-0.13-(0.13 \times I1)-(0.53 \times I4)-(1.30 \times I5)$
	H7	$-0.01-(0.01 \times I1)-(0.01 \times I4)-(0.20 \times I5)$
	H8	$-0.01-(0.01 \times I1)-(0.01 \times I4)-(0.20 \times I5)$
	H9	$-0.23+(1.08 \times I1)+(0.63 \times I2)+(0.38 \times I4)-(0.34 \times I5)$
	H10	$0.83+(0.80 \times I1)+(0.38 \times I2)+(2.21 \times I4)-(0.26 \times I5)$

Table 5.10은 인공신경망 알고리즘 적용 결과 은닉층, 출력층에 가중치를 더하여 도출한 판별함수이다. 최종적으로 O1을 구하는 공식에 따라 독립변수인 Velocity의 Over10, Under10 항목을 구분하게 된다.

인공신경망 판별함수를 통하여 분류된 데이터는 Table 5.11과 같다. 전체 학습데이터 중 25개의 오류가 발생하여 17.99%의 오류율을 가지며, 전체 10cm/s 이상 데이터는 35개에서 18개의 오분류를 하였고 10cm/s 이하 데이터는 104개의 데이터 중 7개만을 오분류 하였다.

Table 5.11 Classified Train data in Artificial Neural Network

Train \ Classified	Over 10	Under 10
Over 10	17	18
Under 10	7	97

인공신경망 결과에 따른 비선형 함수모형에서 독립변수와 종속변수 간 중요성, 즉 접안속도에 영향을 미치는 변수의 순서는 가중치를 이용하여 측정할 수 있다. 독립변수의 상대적 중요도는 가중치 분할법(weight partitioning method)을 이용하여 구할 수 있는데 이는 은닉층과 출력층 사이의 연결가중치를 각각의 독립변수에 대하여 분할방법이다(Garson, 1991). 가중치 분할법에 따른 독립변수 중요도는 Fig. 5.6과 같다.

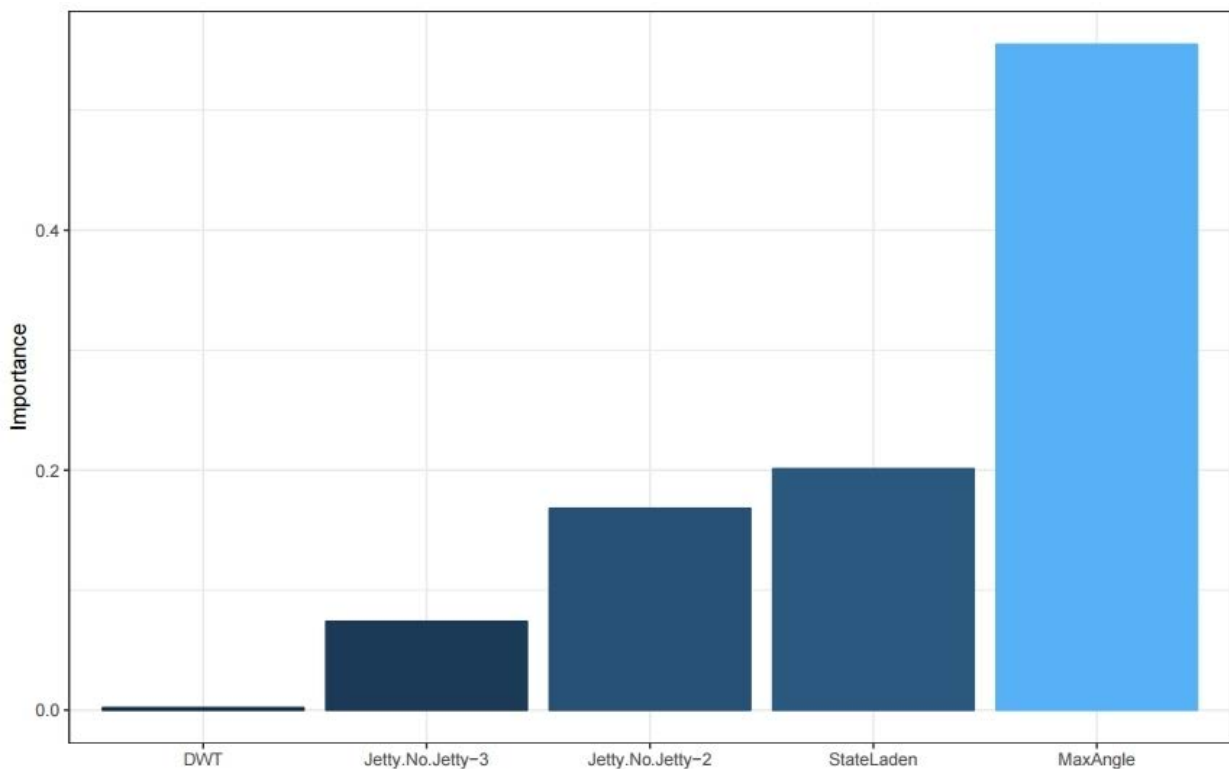


Fig. 5.6 Importance of Independent variable in Artificial Neural Network

5.3 예측모델 성능비교

5.3.1 혼동 행렬

머신러닝 기법에서 모델 결과 평가를 위해 일반적으로 사용하는 방법은 혼동 행렬(Confusion Matrix)을 이용한 평가 지표 계산이다. 혼동 행렬은 머신러닝 모델에서 구한 분류의 예측 범주와 데이터의 실제 분류 범주를 교차표(cross table) 형태로 정리한 행렬을 의미한다. 독립변수, 즉 목표변수가 2-범주인 경우에는 Table 5.12과 같이 표현된다(박인근, 2019).

Table 5.12 2*2 Confusion Matrix

Actual \ Prediction	Y	N
	Y	O (TP : True Positive)
N	X (FP : False Positive)	O (TN : True Negative)

Table 5.12에서 O에 해당하는 경우가 머신러닝 모델이 올바르게 예측한 경우, 즉 예측 범주값이 실제 범주값과 일치한 경우를 말한다. 반대로 X의 경우에는 머신러닝 모델이 부정확하게 예측한 경우를 말하는 것이다. 분석가가 주로 관심있는 범주를 긍정(Positive) 범주라 하고, 그 반대의 범주를 부정(Negative) 범주라 한다. 관심 범주를 정확히 분류한 경우 TP(True Positive), TN(True Negative)이며, 실제로는 관심 범주 이외의 범주임에도 관심 범주로 구분한 경우를 FP(False Positive), 실제로는 관심 범주임에도 관심 외 범주로 구분한 경우를 FN(False Negative)라 한다.

앞서 접안속도 예측모델에서도 Train data set을 이용한 머신러닝 알고리즘 출력결과에 따른 혼동 행렬을 제시하여 오류률에 대하여 언급하였으나, 접안속도 예측모델 성능 비교를 위하여 Test data set(62개)을 입력하여 다시 혼동 행렬을 구하였다(Table 5.13).

Table 5.13 Result of Confusion Matrix

Decision Tree		
Classified \ Test data	Over 10	Under 10
Over 10	9	6
Under 10	14	38

Random Forest		
Classified \ Test data	Over 10	Under 10
Over 10	9	6
Under 10	11	41

Logistic Regression		
Classified \ Test data	Over 10	Under 10
Over 10	8	7
Under 10	11	41

Artificial Neural Network		
Classified \ Test data	Over 10	Under 10
Over 10	7	8
Under 10	6	46

혼동 행렬을 통하여 계산될 수 있는 머신러닝 모델의 평가 지표는 정확도 (Accuracy), 오류율(Error rate, 1-Accuracy), 민감도(Sensitivity), 특이도 (Specificity), 정밀도(Precision)가 있다. 정확도는 실제 분류 범주를 정확하게 예측한 비율을 말하며 오류율은 그 반대로 실제 분류 범주를 잘못 분류한 비율이다. 민감도는 실제로 긍정(Positive)인 범주 중에서 ‘긍정’으로 올바르게 예측 (True Positive)한 비율을 말하며 hit ratio, 재현율(recall), TP rate라고도 한다. 특이도는 실제 부정(Negative)인 범주 중에서 ‘부정’으로 올바르게 예측(True Negative)한 비율을 말한다. 정밀도는 모델이 예측한 데이터 중 올바르게 예측된 데이터의 비율을 말한다.

Table 5.14 Evaluation Index of Machine Learning Model

	Calculation	Decision Tree	Random Forest	Logistic Regression	Artificial Neural Network
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$	0.7015	0.7463	0.7313	0.7910
Error rate	$(FP+FN)/(TP+TN+FP+FN)$	0.2985	0.2537	0.2687	0.2090
Sensitivity	$(TP)/(TP+FN)$	0.6000	0.6000	0.5333	0.4667
Sensitivity	$(TP)/(TP+FN)$	0.6000	0.6000	0.5333	0.4667
Precision	$(TP)/(TP+FP)$	0.3913	0.4500	0.4210	0.5384

Table 5.14는 혼동 행렬을 통하여 계산된 모델 평가 지표의 계산법과 Table 5.13의 결과를 바탕으로 계산한 모델별 평가 지표의 비교를 나타낸 것이다. 단

순히 예측모델의 정확도를 비교할 수 있는 Accuracy의 경우 인공신경망, 랜덤 포레스트, 로지스틱회귀, 의사결정나무 순으로 나타났다. 하지만 정확도는 실제 예측하고자 하는 데이터를 제대로 예측하지 못하더라도 높게 나올 수 있기 때문에 모델을 평가하는 지표로 사용하기에는 한계가 있다. 따라서 혼동 행렬을 통해 모델의 성능을 평가하기 위해서는 민감도와 정밀도를 이용한 식(5.1)을 이용한다(Olson et al., 2008).

$$F\text{-measure} = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (5.1)$$

식 5.1을 계산하기 쉽게 정리하면 식 5.2와 같다.

$$F\text{-measure} = \frac{2TP}{2TP + FN + FP} \quad (5.2)$$

식 5.2에 따라 혼동 행렬을 통한 예측모델 성과를 수치화하여 Table 5.15와 같이 정리하였다.

Table 5.15 F-measure

	Algorithm			
	Decision Tree	Random Forest	Logistic Regression	Artificial Neural Network
F-measure	0.4737	0.5143	0.4706	0.5000

따라서 혼동행렬을 통한 평가지수에 따라 머신러닝 예측모델의 성능은 랜덤 포레스트, 인공신경망, 의사결정나무, 로지스틱회귀 순으로 나타났다.

5.3.2 ROC 곡선

ROC 곡선(Receiver Operation Characteristic curve, ROC curve)은 참 긍정비율(TP rate)과 거짓 긍정비율(FP rate)의 관계를 나타낸 그래프이다(Fig. 5.7). 참 긍정비율은 민감도와 같은 개념이며, 거짓 긍정비율은 실제 부정인 범주 중에서 긍정으로 잘못 예측한 비율로서 1-특이도와 같은 개념이다.

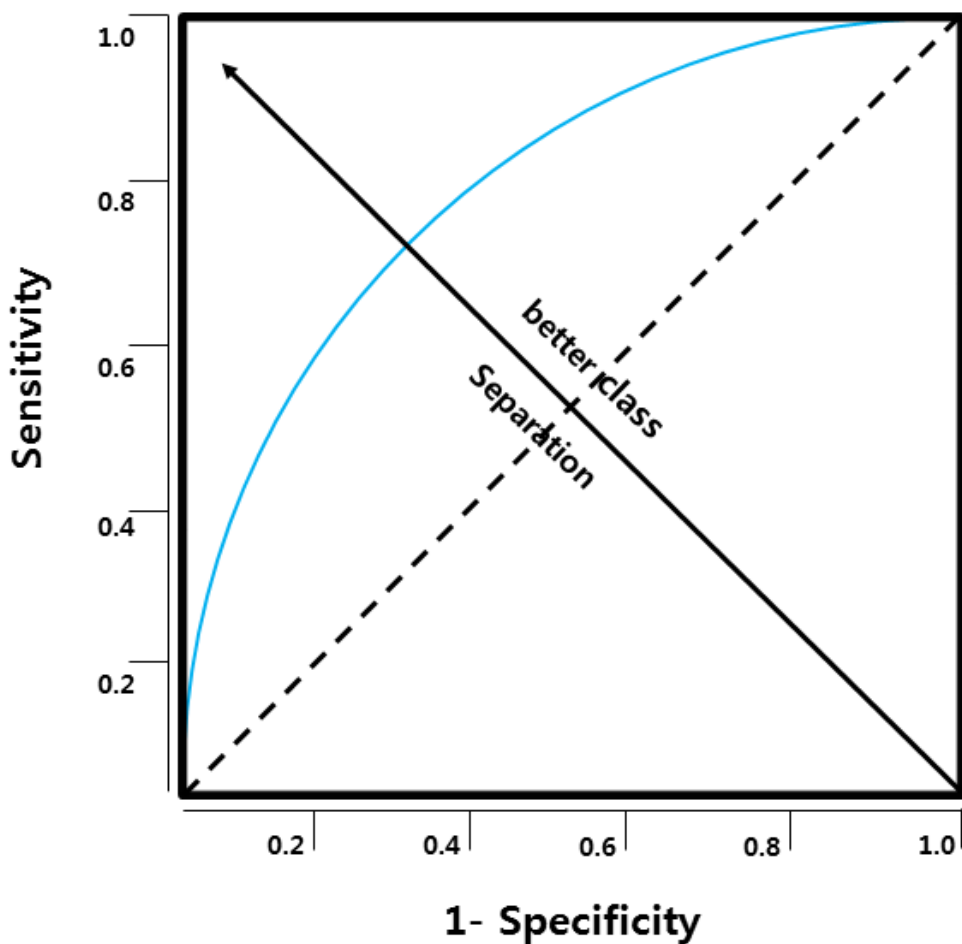


Fig. 5.7 ROC curve

ROC 곡선의 밑 면적을 AUC(the Area Under a ROC curve)라고 하는데 AUC가 1에 가까울수록, 즉 ROC 곡선이 왼쪽 상단 꼭지점에 다가갈수록 좋은 성능을 가진다고 할 수 있다. 이 때 AUC는 0.5 ~ 1사이의 범위를 가진다. AUC는 전체적인 민감도와 특이도의 상관관계를 보여줄 수 있어 대표적인 모델 성능 측정 기준이 된다.

본 연구의 접안속도 예측모델 구축에 따른 ROC 곡선은 Fig 5.8과 같다. ROC 곡선의 AUC값에 따라 랜덤포레스트, 인공신경망, 의사결정나무, 로지스틱회귀 모델 순으로 예측모델 성능이 나타난다.

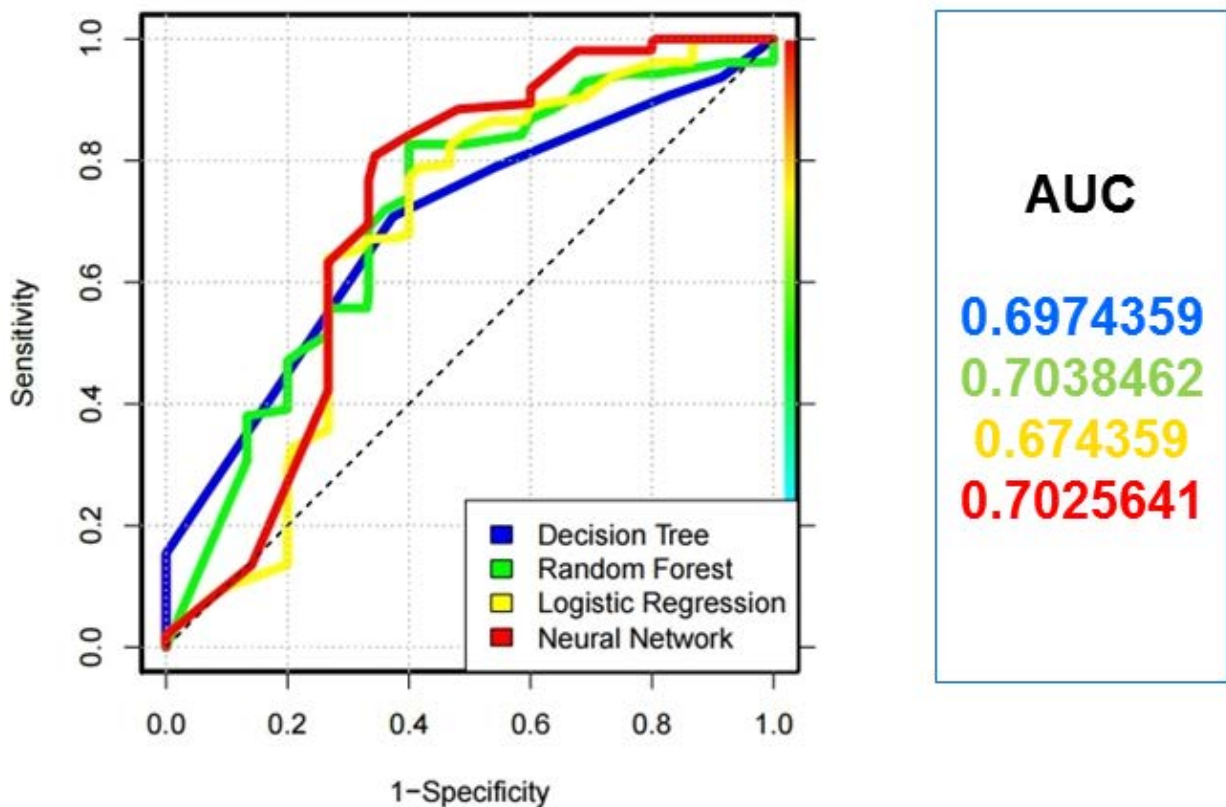


Fig. 5.8 Compared with ROC curve

5.4 최종모델 선택 및 영향요인 분석

혼동행렬과 ROC 곡선에서 예측모델 알고리즘의 성능을 평가하여 최종 접안 속도 예측모델을 채택하고 분석하기 위하여 다음과 같은 사항을 고려하였다.

첫째, 선박의 접안속도 예측모델의 중요한 사항은 10cm/s 이상의 접안 속도를 가진 선박을 예측하는 것이다. 이는 본 연구의 접안속도 예측모델을 구축하는 이유이며 선박 접안 시, 10cm/s 이상의 위험한 접안속도를 가지는 경우를 예측하여 사고에 대비하기 위함이다.

둘째, 혼동행렬 및 ROC 곡선과 같은 성능평가를 통해 우수한 성능을 가진 알고리즘을 선별하여 접안속도 예측모델로 최종 채택 한다.

셋째, 최종 채택한 접안속도 예측모델의 종속 변수의 중요도에 따라 접안속도에 영향을 미치는 요소를 판별한다.

이에 따라 혼동행렬과 ROC 곡선을 이용하여 접안속도 예측모델의 성능을 최종 비교한 결과를 Table 5.16로 나타내었다. F-measure 값과 ROC 곡선의 AUC 값을 포함하였다. 4개의 알고리즘 별 지표의 순위에 따라 1에서 4의 값을 책정하고 가장 높은 순위를 보인 알고리즘을 접안속도 예측모델로 채택한다.

Table 5.16 Adopting the final machine learning model

	Algorithm			
	Decision Tree	Random Forest	Logistic Regression	Artificial Neural Network
F-measure	0.4737 (3)	0.5143 (1)	0.4706 (4)	0.5000 (2)
AUC	0.6974 (3)	0.7038 (1)	0.6744 (4)	0.7026 (2)
Adoption		√		

접안속도 예측모델의 최종 평가 결과 혼동행렬을 통하여 확인할 수 있는 지표인 F-measure는 0.5143으로 랜덤포레스트가 가장 좋은 성능을 보였고, ROC 곡선에서 또한 랜덤포레스트가 AUC값이 0.7038462로 가장 좋은 성능을 보였다. 따라서 성능 평가 결과 최종 접안속도 예측모델은 랜덤포레스트로 채택하였다.

랜덤포레스트를 최종 접안속도 예측모델로 채택함에 따라 접안속도에 영향을 미치는 요소는 Fig. 5.4의 정확도 감소성으로 판별할 수 있다. 랜덤포레스트 예측모델의 정확도 감소성에 의한 접안속도에 가장 영향을 미치는 요소는 STATE로서 선박의 접안 시 선박운항자는 화물의 적재 상태에 따라 접안속도에 영향을 가장 많이 받았다는 것을 알 수 있다. STATE 다음의 중요한 요소로는 DWT, Jetty No., MaxAngle 순이다.

접안속도에 가장 영향을 미치는 요인이 STATE임에 따라 Ballast(66.0%) 상태인 선박이 상대적으로 접안속도 10cm/s 이상으로 접안한 것을 알 수 있다 (Table 4.8). 따라서 선박이 접안 할 때 Ballast 상태인 경우 선박운항자는 더욱 주의하여 접안하는 것이 사고 예방에 도움이 될 것이다.

제 6 장 결론 및 향후과제

6.1 결론 및 향후과제

본 논문에서는 접안속도 측정 장치가 설치되어있는 국내항 중 Jetty 1~3으로 구분된 탱커 터미널 부두에서 수집한 접안속도 실측 데이터를 바탕으로 머신러닝 기법을 이용하여 예측모델을 구축하여 성능을 비교하고 영향을 미치는 요인을 분석하였다. 선박의 접안 시 접안에너지가 발생하며 접안에너지가 과도한 경우 부두 및 선박의 손상으로 이어지게 된다. 접안속도는 접안에너지를 산정하는 요소 중 가장 영향을 크게 미치는 요소로서 선박의 접안 시 적정 접안속도를 유지하여 접안하는 것이 중요하다. 따라서 접안속도 예측모델을 구축하여 접안속도를 예측하고 접안속도에 영향을 미치는 요인을 파악할 수 있다면 선박운항자 및 부두관리자 입장에서 사고 대비 및 대응이 가능하다.

본 연구에서는 2017년 3월부터 2018년 7월까지 약 17개월 간 대상 부두에 접안한 선박에서 측정한 데이터를 취합하였다. 또한 기상청의 자료를 이용해 해당 기간의 날씨 데이터를 수집하였다. 선박의 규모는 약 10,000 DWT에서 320,000 DWT까지 다양하며 선종은 탱커에 한정되어있다. 수집된 데이터를 바탕으로 접안속도 예측모델 구축에 필요한 독립변수를 PIANC WG145(2017)를 참조하여 채택하였으며, 결측값은 몬테카를로 방법으로 처리하였다.

접안속도 예측모델 분석에 있어 선정된 독립변수인 Jetty No., DWT, STATE, MaxAngle, Pilot, Wind를 연관성 분석인 상관분석 및 교차분석을 통하여 최종 독립변수로 Jetty No., DWT, STATE, MaxAngle을 선정하였다. 선정된 변수를 바탕으로 다중선형회귀분석을 통하여 선박 접안속도 예측식을 도출하고자 하였으나 R^2 값이 0.132으로 예측모델의 성능이 현저히 떨어져 머신러닝 기법을 활용하여 예측모델을 구축하였다.

머신러닝 알고리즘으로는 의사결정나무, 랜덤포레스트, 로지스틱회귀, 인공신경망을 활용하였다. 예측모델의 성능비교를 위하여 홀드아웃 방식을 채택하여

데이터를 Train data set과 Test data set으로 분리하여 분석을 진행하였다.

머신러닝 알고리즘을 적용하여 분석결과 의사결정나무, 랜덤포레스트, 로지스틱회귀, 인공신경망의 F-measure 값은 각각 0.4737, 0.5143, 0.4706, 0.5000으로 랜덤포레스트가 가장 높았으며, ROC 곡선을 이용한 성능평가에서도 AUC값이 각각 0.6974, 0.7038, 0.6744, 0.7026로 랜덤포레스트가 가장 좋은 성능을 보였다.

따라서 F-measure 값과 ROC 곡선을 통하여 좋은 성능을 보인 랜덤포레스트를 최종 예측모델로 채택하였다. 랜덤포레스트에 따른 접안속도에 영향을 미치는 요인은 STATE, DWT, Jetty No., MaxAngle 순으로 나타났다.

본 논문은 선박 운용분야에서 빅데이터에 기반한 머신러닝 기법을 이용한 연구사례로서 선박의 접안속도 예측모델을 구축하였다. 또한 선박 운용분야에서도 데이터 축적을 통하여 빅데이터에 기반한 머신러닝, 인공지능 등의 기법으로 예측모델, 의사결정모델 등의 활용 가능성을 제시하였다.

하지만 이러한 연구결과에도 본 연구는 다음과 같은 한계점을 가지고 있다.

첫째, 다양한 선종 및 부두의 데이터 수집이 필요하다. 본 연구에서는 한 탱커부두에서 수집된 데이터와 그에 따른 탱커선의 접안 데이터에 한정된 연구를 진행하였다. 선박은 다양한 선종이 있으며, 부두 특성 또한 선종별로 상이하므로 선박의 접안속도 예측에 관한 연구 진행 시, 다양한 선종 및 부두의 데이터가 필요한 것으로 사료된다.

둘째, PIANC에서 제시한 접안속도에 미치는 영향 요인 모두를 분석에 활용하지 못하였다. 날씨 데이터 중 Current 및 Wave, 도선사, 예인선 등의 자료를 수집하지 못하여 분석에 활용하지 못하였는데 이와 같은 데이터를 수집하여 분석 자료로 활용한다면 더욱 정확한 예측모델이 구축될 것이다.

셋째, 데이터양의 한계점이다. 본 연구에서는 약 17개월간 206개의 데이터를 활용하였으나 긴 기간동안 많은 데이터를 수집한다면 더욱 정확한 예측모델의 구축이 가능하다.

이러한 한계점을 보완하여 후속 연구가 진행된다면 더욱 정확하고 높은 성능을 가진 접안속도 예측모델이 구축될 것이라 생각된다. 정확한 접안속도 예측모델이 구축된다면

항만 관리자, 선박 운항자의 안전관리 측면에서 큰 도움이 될 것이다.

앞으로 빅데이터 분석, 머신러닝, 인공지능 등의 기법은 선박 및 해운분야에서 중요한 역할을 수행할 것이다. 선박 센서 및 항만 부두에서 수집된 정보를 통해 선박 운항의 안전성만이 아니라 선박 및 부유지에도 두 설계 및 안전성 개선이 될 것이다. 더하여 4차산업혁명 및 자율운항선박 시대가 도래함에 따라 빅데이터 분석, 머신러닝, 인공지능 등의 활용은 필수불가결할 것이며 많은 연구가 필요하다.

참고문헌

- [1] 강동훈, 2018. *머신러닝 기법을 이용한 너울성 파랑 발생 예측에 관한 연구*. 석사학위논문. 서울: 단국대학교.
- [2] 김병수, 배화수, 석경하, 조대현, 최국렬, 2018. *SAS EM 14.1 데이터마이닝 기초와 활용*. 서울: (주)교우.
- [3] 김지혜, 2018. *조선 생산 리드타임 예측을 위한 기계학습 방법론에 관한 연구*. 석사학위논문. 부산: 한국해양대학교.
- [4] 김진호, 2016. *빅데이터가 만드는 제4차 산업혁명: 개인과 기업은 어떻게 대응할 것인가?* 서울: 북카라반.
- [5] 김호영 2005. *고속도로의 속도예측을 통한 안전성평가 모형 개발*. 박사학위논문. 서울: 연세대학교.
- [6] 남규민, 2018. *데이터마이닝 기법을 이용한 기업부실 예측모형의 성과 비교*. 석사학위논문. 부산: 부산대학교.
- [7] 박선우, 2018. 빅데이터 시대와 데이터 융합. *정보통신방송정책*, 제30권 1호 통권 661호.
- [8] 박승수, 2016. *온라인 리뷰와 머신러닝을 활용한 드라마 시청률 예측 모델 연구*. 석사학위논문. 서울: 연세대학교.
- [9] 박인근, 홍지후, 강남규, 김성호, 정구범, 2019. *4차 산업혁명 현장 전문가가 알려주는 빅데이터 분석과 활용*. 파주: (주)제이펍.
- [10] 이상원, 2019. *선박의 안전한 접안을 위한 실측 접안속도 분석에 관한 연구*. 석사학위논문. 부산: 한국해양대학교.
- [11] 이정진, 2011. *R, SAS, MS-SQL을 활용한 데이터마이닝*. 파주: 자유아카데미.
- [12] 이훈영, 2008. *이훈영교수의 연구조사방법론*. 서울: 도서출판 청람.

- [13] 임상섭, 2018. *Development of Machine Learning Methods to Improve Decisions in Bulk Chartering Practice*. 박사학위논문. 부산: 한국해양대학교.
- [14] 전해남, 2018. 융합보안과 인공지능 - 머신러닝, 딥러닝. *SK infosec 공식블로그* (<http://blog.skinfosec.com/221396731850>)
- [15] 조국연 2018. *화주의 내륙운송 의사결정요인과 영향에 대한 연구(한국과 북중국을 중심으로)*. 박사학위논문. 인천 : 인천대학교.
- [16] 조익순, 조장원, 이상원, 2018. 선박 접안속도의 실측데이터 분석에 관한 기초 연구. *한국연안방재학회지*, 5(2), pp.105~112.
- [17] 최병근, 황용근, 정교민, 2014. 소셜 네트워크 정보 확산 및 구조 분석을 위한 머신러닝 및 데이터 마이닝 기법. *정보과학회지*, pp21~25.
- [18] 함유근, 채승병, 2012. *빅데이터, 경영을 바꾼다*. 서울: 삼성경제연구소
- [19] 해양수산부, 2017a. *항만 및 어항설계기준(상)*, 세종:해양수산부.
- [20] 해양수산부, 2017b. *항만 및 어항설계기준(하)*, 세종:해양수산부.
- [21] Brett Lantz, 2014. *R을 활용한 기계학습*. 서울: (주)에이콘출판.
- [22] Kevin P. Murphy, 2016. *머신러닝 Machine Learning*. 서울: (주)에이콘출판.
- [23] Raghav Bali, Dipanjan Sarkar, 2017. *R예제로 배우는 머신 러닝, 머신 러닝의 기본 이해와 실생활 문제 해결에의 적용*. 서울: (주)에이콘출판.
- [24] Beckett Rankine, 2010. Berthing velocities and Brolsma's curve. *BR Paper Review of Berthing Condition Definitions-Rev 1.doc*
- [25] Alfred Roubos, Leon Groenewegen, Dirk Jan Peters, 2017. Berthing velocity of large seagoing vessels in the port of Rotterdam. *Article in Marine Structures*.
- [26] Bridgestone, 2011. *Marine fender catalogue*. Tokyo: Bridgestone.
- [27] British Standard Institute, 2010. *BS 6349-2:2010 Maritime Works-Part 2 : Code of practice for the design of quay walls, jetties and dolphins*. London: British Standard

Institute.

[28] British Standard Institute, 2014. *BS 6349-4:2014 Maritime Works-Part 4 : Code of practice for design of fendering and mooring systems*. London: British Standard Institute.

[29] Broolsma, J.U., 1977. On Fender Design and Berthing Velocities. *Proc. International Navigation Congress*, Section II, Subject 4, pp.87-100.

[30] Caser Wu, Rajkumar Buyya, Kotagiri Ramamohanarao, 2016. Chapter 1 Big Data Analysis = Machine Learning + Cloud Computing. *Big Data: Principles and Paradigms. Elsevier, Inc.*

[31] Costa, V., 1964. The berthing ship-The effect of impact on the design of fenders and berthing structures. *The Dock and Harbour Authority*, Vol.45.

[32] Gason, 1991. Interpreting Neural Network Connection Weights. *AI Expert 6(7)*, pp47-51.

[33] <http://www.microsoft.com/bigdata>. *Microsoft*

[34] Jacob Cohen, 1969. *Statistical power analysis for the behavioral sciences*-second edition. *LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS*.

[35] Japanese technical standard, OC DI, 2009. *Technical Standards and Commentaries for Port and Harbour Facilities in Japan*. Tokyo, Japan. *The Overseas Coastal Area Development Institute*.

[36] J. Gantz, D. Reinsel, 2012. *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*. *IDC iView*.

[37] Laney D., 2001. 3D data management: controlling data volume, velocity and variety. *USA: Application Delivery Strategies*, Meta Group; 2001. p. 1-4.

[38] Mario Milicevic, Ines Obradovic, 2018. Big Data in the Maritime Industry. *Nase More*, 65(1), pp56~62.

- [39] Olson, David L.; and Delen, Dursun, 2008. *Advanced Data Mining Techniques*, Springer, 1st edition (February 1, 2008), page 138, ISBN 3-540-76916-1
- [39] Ornulf Jan Rodseth, Lokukaluge Prasad Perera, Brage Mo, 2016. Big data in shipping – Challenges and opportunities. In *Proceedings of the 15th International Conference on Computer and IT Applications in the Maritime Industries (COMPIT 2016)*, Lecce, Italy, May 2016.
- [40] PIANC, 2002. Guidelines for the design of fenders systems, *Report of working group 33 of the MARITIME NAVIGATION COMMISSION*, Brussel: PIANC.
- [41] PIANC, 2017. Berthing velocities and fender design, *Draft Report of working group 145 of the MARITIME NAVIGATION COMMISSION*, Brussel: PIANC.
- [42] Roubos, A., Groenewegen, L. & Peters, D.J., 2016. Berthing velocity of large seagoing vessels in the port of Rotterdam. *Marine Structures*, 51, pp.202-219
- [43] Sadaharu Koga, 2015. *Major Challenges and solutions for utilizing big data in the maritime industry*. A master's thesis. Malmö: World Maritime University.
- [44] Samuel A., 1959. Some studies in machine learning using the game of checkers. *IBM J Res Dev 1959*; 3(3): 211-29.
- [45] Sonia Singh, Priyanka Gupta, 2014. Comparative Study ID3, CART and C4.5 Decision Tree Algorithm: A Survey. *International Journal of Advanced Information Science and Technology(UAIST) Vol.27*, No27.
- [46] *Spanish ROM 3.1-99*
- [47] S. Ueda, 1981. Study on berthing impact force of very large crude oil carriers, Report of Port and Harbour Research Institute.
- [48] , S. Ueda, Horii O., K. Ichikawa, 1975. Investigation on Crude Oil Tanker Berth, *Technical Note of the Port and Harbour Research Institute*, No.201.
- [49] TRELLEBORG, Use of big data in the maritime industry. *Trelleborg Marine*

Systems and Port Technology.