Journal of Next-generation Convergence Information Services Technology Vol.10, No.1, February (2021), pp.1-9
ISSN 2384-101X(p) 2672-1163(e) *http://dx.doi.org/10.29056/jncist.2021.02.01*

합성곱신경망을 이용한 타이타닉호의 재난 데이터 분석

Analysis for the RMS Titanic Disaster Data Using Convolutional Neural Networks

이석준[1], 심동희[2]*

Seokjun Lee[1], Donghee Shim[2]*

요 약

본 연구에서는 합성곱신경망을 이용해서 타이타닉 재난데이터를 분석하였다. 그 동안 타이타닉 재난 데이터에 대한 k-최근접이웃 알고리즘, Naive Bayes 알고리즘, 의사결정트리 알고리즘 , SVM 알고리즘, 로지스틱 회귀 알고리즘 등 인공지능기법을 이용한 많은 분석이 이루어져왔다. 또한 인공신경망을 이용한 분석도 이루어져왔지만 합성곱신경망을 이용해서 이 타이타닉 재난데이터에 대한 분석은 없었다. 이 타이타닉 재난데이터 분석을 위하여 Keras를 이용하여 합성곱신경망중 LeNet-5를 사용해 모델링하여 훈련을 진행하였다. 이 성능결과를 Kaggle 웹사이트에 올려진 다중신경망을 이용한 분석결과와 비교하였다. LeNet-5를 이용한 경우 모델훈련에서 요구되는 훈련파라미터 개수의 비교에서 Kaggle 웹사이트에 올려진 다중신경망보다 월등히 적었다. 또한 훈련된 모델을 이용한 테스트 결과에서도 정확도는 0.856, 정밀도는 0.924를 나타냈다. 이 성능측정치는 Kaggle에 올려진 다중신경망의 정확도 0.842, 정밀도 0.873보다 우수한 것이다.

핵심어 : 타이타닉 재난데이터, 합성곱신경망, LeNet-5, Keras

Abstract

Titanic disaster data is analyzed using the convolutional neural networks in this paper. Although many researchers have analyzed this data using the aritificial intelligence techniques such as k-nearest neighbors algorithm, Naive Bayes algorithm, decision tree algorithm, SVM algorithm and logistic regression algorithm. Although the artificial neural networks are also used in another researches, the convolutional neural networks model has not been used in the analysis for this titanic disaster data. LeNet-5 among the various convolutional neural networks model is used with Keras for the train in this analysis. The performance result is compared with the result of analysis using the mutlilayer neural networks uploaded in Kaggle web site. In the comparison of the number of parameters to be trained, LeNet-5 shows efficient than multilayer neural networks. In the results of test after trained process, LeNet-5 shows 0.856 in accuracy and 0.923 in precision. These performance measures are better than those of multilayer neural networks whose accuracy was 0.841 and the precision was 0.872.

Keyword : Titanic Disaster Data, Convolutional Neural Networks, LeNet-5, Keras

1 Department of Carbon Convergence Engineering, Jeonju University, Jeonju, Korea [Graduate Student]
e-mail: qbox3d@gmail.com

2 Department of Computer Science & Engineering, Jeonju University, Jeonju, Korea [Professor]
e-mail: dhshim@jj.ac.kr (Corresponding author)

Received(January 11, 2021), Review Result(1st: January 27, 2021), Accepted(February 5, 2021), Published(February 28, 2021)

© 2021 The Authors. Published by NCISS.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/4.0/.

1. 서론

RMS 타이타닉(Titanic)호는 영국 선박회사에서 운영한 북대서양 횡단 여객선이었다 [1]. 1912년 4월 10일 영국의 사우샘프턴을 출발해서 미국의 뉴욕으로 향하던 첫 항해중 4월 15일에 빙산과 충돌하여 타이타닉호는 침몰하여 1,514명이 사망한 가슴 아픈 해난사고가 되었다.

이 사고를 시나리오로 한 영화도 제작되어 재난사고에 대한 많은 교훈을 남겨주기도 했다. 또한 이 사고에 대하여 선박의 구조적인 문제에 대한 연구 [2][3], 포렌식 측면에서의 연구 [4] 등이 이루어졌다. 한편 이 사고로 인한 사망자와 생존자에 대한 데이터분석을 인공지능적인 방법을 이용한 연구도 이루어졌다 [5-8]. 그러나 이 인공지능적인 방법에서 인공신경망을 이용한 연구 [5][9]도 드물게 있었지만 일반적인 다중계층신경망을 이용하였으며 합성공신경망(CNN:Convolutional Neural Networks) [10]을 이용한 분석은 없었다.

본 논문에서는 타이타닉 재난의 데이터에 대하여 파이썬을 이용하여 Keras 라이브러리 [11]를 이용하여 합성공신경망에 해당하는 LeNet [10]으로 모델링하여 구축하였다. 데이터를 이용하여 훈련을 거친 후에 성능평가를 하였다. 성능평가결과 다른 인공신경망에서의 성능보다 더 좋은 성능을 나타냈다.

본 연구의 2장에서는 타이타닉재난데이터에 대한 분석에 사용된 기법들을 간략히 소개하고, 3장에서는 데이터 설명, LeNet 모델구축에 대하여 설명하고, 4장에서는 구현과 그 결과를 나타냈으며, 5장에서는 결론을 기술하였다.

2. 타이타닉 재난 데이터에 대한 분석방법

일반적으로 기계학습에서는 데이터검토, 데이터선처리, 학습모델 및 알고리즘선택, 학습수행 및 성능평가의 절차로 이루어진다 [9][10]. 데이터검토에서는 데이터가 분류문제인 지, 범주형인 지 수치형인 지 점검한다. 데이터선처리에서는 특징을 추출하고, 데이터형태의 변경, 누락값 처리, 표준화 등의 작업을 한다. 학습모델과 알고리즘선택에서는 적용할 모델과 알고리즘을 선택한다. 그리고 선택된 모형화를 하고 알고리즘을 이용해서 훈련세트를 이용해서 기계학습을 하고 시험세트를 이용해서 성능평가를 한다.

타이타닉 재난데이터를 인공지능의 여러 모델을 이용한 연구 [5-9]에서는 다양한 분석을 하였다. 여기에서 사용된 알고리즘은 k-최근접이웃(k-Nearest Neighbors) 알고리즘, Naive Bayes 알고리즘, 의사결정트리 알고리즘, 앙상블 알고리즘(랜덤포리스트, xgboost 등), SVM(Support Vector Machine) 알고리즘, 로지스틱 회귀알고리즘, 신경망 알고리즘 등이었다. 이중 가장 많은 방법으로 사용된 것은 랜덤포리스트와 신경망 알고리즘이었다.

k-최근접 이웃(k-Nearest Neighbors) 알고리즘 [12]은 분류나 회귀에 사용되는 비모수 방식이다. 분류나 회귀 모두 입력이 특징 공간내 k개의 가장 가까운 훈련 데이터로 구성되어 있다. 그래서 테스트 데이터셋의 레이블이 없는 각 레코드에 대해 훈련데이터에서 유사도 기준으로 K개의 가장 가까운 값을 찾는다. 출력은 분류에서는 소속된 항목이며, 회귀에서는 객체의 특성 값이다.

Naïve Bayes 알고리즘 [13]은 특성들 사이의 독립을 가정하는 베이즈 정리를 적용한 확률 분류기의 일종으로 가장 높은 확률의 레이블로 분류를 한다.

의사결정 트리 [14]는 의사결정 규칙과 그 결과들을 트리구조로 도식화한 의사결정 지원도구의 일종으로 변수들 간의 규칙, 관계 등으로 레이블을 분류하는 트리구조의 모델을 생성하고, 관측값을 해당 모델에 대입하여 레이블을 예측하는 방식이다.

앙상블 알고리즘 [15]에서는 더 좋은 예측성능을 얻기 위하여 다수의 학습알고리즘을 이용하는 방식이다. 그리하여 샘플을 여러 번 출력하여 각 모델을 학습시켜 결과를 집계하는 bagging 방식(병렬)과 순차적으로 가중치를 부여하는 boosting 방식이 있다.

SVM (Support Vector Machine) 알고리즘 [16]은 두 카테고리 중 어느 하나에 속한 데이터 집합이 주어졌을 때, 주어진 데이터 집합을 바탕으로 하여 새로운 데이터가 어느 카테고리에 속할 지 판단하는 비확률적 이진선형분류 모델을 만든다. 만들어진 분류모델은 데이터가 사상된 공간에서 경계로 표현되는데 그 중 가장 큰 폭을 가진 경계를 찾는 알고리즘이다. 선형분류와 더불어 비선형분류에서도 사용될 수 있다. 비선형분류를 하기 위해서 주어진 데이터를 고차원 특징 공간으로 사상하는 작업이 필요한데, 이를 효율적으로 하기 위해 커널 트릭을 사용하기도 한다.

로지스틱 회귀알고리즘 [17]은 독립변수들의 선형결합을 이용하여 사건의 발생 가능성을 예측하는데, 일반적인 회귀분석의 목표와 동일하게 종속변수와 독립변수간의 관계를 구체적인 함수로 나타내 향후 예측모델에 사용하는 것이다. 이는 독립변수의 선형 결합으로 종속변수를 설명한다는 관점에서는 선형 회귀분석과 유사하다. 하지만 로지스틱 회귀분석은 선형 회귀분석과는 다르게 종속 변수가 범주형 데이터를 대상으로 하며 입력 데이터가 주어졌을 때 해당 데이터의 결과가 특정 분류로 나뉘기 때문에 일종의 분류기법으로도 볼 수 있다.

인공신경망 [18]은 시냅스의 결합으로 네트워크를 형성한 노드들이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제해결 능력을 가지는 모델 전반을 가리킨다. 각 층은 이러한 노드들로 구성되는 데, 층은 입력층과 중간의 은닉층, 출력층으로 구성된다. 인공신경망은 좁은 의미에서는 오차역전파법을 이용한 다층 퍼셉트론을 가리키는 경우가 많다. 인공신경망에는 정답에 해당하는 신호의 입력에 의해서 문제에 최적화되어 가는 지도학습과 정답에 해당하는 신호를 필요로 하지 않는 비지도학습이 있다. 명확한 해답이 있는 경우에는 지도학습이, 데이터 클러스터링에는 비지도학습이 이용된다. 인공신경망은 많은 입력들에 의존하면서 일반적으로 알 수 없는 함수를 추측하고 근사치를 낼 경우 사용한다. 일반적으로 입력으로부터 값을 계산하는 뉴런시스템의 상호연결로 표현되

고 적응성이 있어 패턴인식과 같은 기계학습을 수행할 수 있다.

[5]번 연구에서는 캐글(Kaggle) [19]에 올려진 튜토리얼모델 5개를 비교분석하면서 확장적인 튜토리얼모델을 개발하였다. 캐글은 사용자가 해결과제와 데이터를 먼저 등록하면, 이에 대한 해결책으로 다른 사용자들이 예측모델을 개발해서 올림으로써 경쟁하는 플랫폼이다. 경쟁의 경우 도출하고자 하는 결과에 대한 정보와 이를 요구하는 포맷에 맞게 등록하면 결과에 따른 점수를 확인할 수 있다. 반면에 데이터셋은 참여자들이 자유롭게 데이터를 공유하기 위한 것이다. 이 연구에서 튜토리얼모델 1번부터 4번까지는 모두 랜덤포리스트 알고리즘의 정확도가 제일 높이 나타났다. 튜토리얼모델 5번은 xgboost만으로 분석을 하였는 데 이 정확도는 88.7%를 나타내 다른 튜토리얼모델에서 사용된 다른 알고리즘들보다 높았다. 한편 이 연구에서 구성한 튜토리얼모델 6번에서는 더 다양한 알고리즘을 수용하여 구현해주었으며 인공신경망은 75.5%의 정확도를 나타냈다.

6번 연구에서는 로지스틱 회귀분석, 의사결정트리, k-최근접 이웃, SVM 알고리즘을 이용해서 분석했다. 그 결과 정확도는 79.8%, 91.7%, 86.5%, 79.1%를 각각 나타냈다.

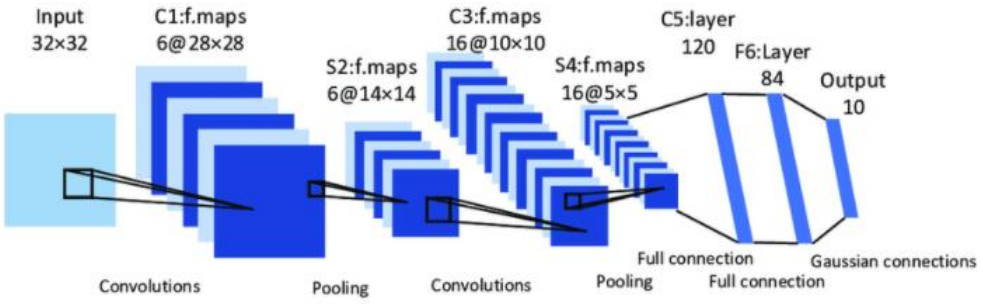
한편 8번 연구에서는 Naive Bayes, 로지스틱 회귀분석, 의사결정트리, 랜덤포리스트 알고리즘을 이용해서 분석했으며 그 결과 정확도는 각각 91.3%, 93.5%, 93.1%, 91.9%를 나타냈다.

이와 같은 연구결과 같은 알고리즘을 이용해도 평가에 사용된 데이터셋이 다소 다르며, 각 데이터의 처리방식에서 다소의 차이에 기인한 것으로 판단한다.

3. 타이타닉 재난데이터분석을 위한 LeNet-5 모델과 데이터구성

3.1 모델의 구성

LeNet은 합성곱신경망을 처음으로 개발한 Yann Lecun 연구팀이 1998년에 개발했다 [11]. LeNet의 여러 버전중 LeNet-5는 [그림 1]에 나타낸 바와 같이 입력층, 3개의 컨볼루션층(C1, C2, C3), 2개의 서브샘플링층(S2, S4), 1개의 완전연결층(F6) 그리고 출력층으로 구성되어 있으며 입력층, C1, S2, C3, S4, C5, F6, 출력층의 순서로 구성된다. LeNet-5 최초의 표준구성에서 C1에서는 32*32 이미지를 입력받아서 6개의 5*5 필터와 컨볼루션 연산을 해서 6장의 28*28 특성맵을 얻는다. S2 에서는 6장의 28*28 특성맵에 대해 서브샘플링을 진행한다. 그래서 14*14의 특성맵으로 축소된다. C3에서는 6장의 14*14 특성맵에 컨볼루션 연산을 수행해서 16장의 10*10 특성맵을 산출한다. S4 에서는 16장의 10*10 특성앱에 대해서 서브샘플링을 진행해 16장의 5*5 특성맵으로 축소한다. C5 에서는 16장의 5 x 5 특성맵을 120개 5 x 5 x 16 사이즈의 필터와 컨볼루션해 120개 1 x 1 특성맵이 산출된다. F6 에서는 84개의 유닛을 가진 피드포워드 신경망인데 . C5의 결과를 84개의 유닛에 연결시킨다. 출력층은 F6의 84개 유닛으로부터 입력을 받아서 10개의 Euclidean radial basis function(RBF) 유닛들로 구성되어 최종적으로 해당 클래스를 알려준다.



[그림 1] LeNet-5의 구성

[Fig. 1] Configuration of LeNet-5

3.2 재난데이터의 구성

재난데이터는 총 11개의 항목으로 구성된 1,309개의 레코드인 데, 훈련용데이터셋은 891개, 테스트용데이터셋은 418개로 나뉘어져있다 [9]. Survival은 사망시에 0, 생존시에 1로 표시되는 종속변수인 바 이 재난데이터는 지도학습으로 처리해야 한다. 항목구성은 [표 1]에 나타난 바와 같다.

[표 1] 타이타닉 재난데이터의 항목 구성

[Table 1] The Fields of Titanic Disater Data

항목	데이타타입	값	의미	결측 레코드수, 값의 종류수
Survival	int64	0=no, 1=yes	survive 여부	0
Pclass	int64		승객등급	0, 3
Name	object	Title생성 사용	특성분류가능	
Sex	int64	0=Male, 1=Female	남성, 여성	0
Age	float64	1-5구간 변형	나이	263 (177, 86)
SibSp	int64	alone생성 사용	같이 승선한 형제자매(부부)수	0
Parch	int64	alone생성 사용	같이 승선한 부모(자녀)수	0
Ticket	object	미사용	티켓번호	0
Fare	float64	1-4구간 변형	요금	1 (0,1), 248
Cabin	object	미사용	선실구분	687, 147
Embarked	object	QSC를 0,1,2로	출발항구	2(2,0), 3
PassengerID	int64	미사용		
Alone	int64/생성	1=단독여행 0=No	SibSp, Parch에서 생성	
Title	int64/생성	Mr, Miss, Mrs, Master, 기타	Name에서 생성	

Pclass는 승객의 사회적인 등급을 나타내며 3가지 경우로 구성되어 있다. 이름은 문자열인데 여

기에 Mr, Miss, Mrs, Master, 기타 정보가 포함되어 있어서 이 특징을 추출하여 Title이라는 항목을 생성하였다. Sex는 남성은 1로 여성은 1로 변환하였다. Age는 5구간으로 나누어 1-5로 변환하였다. SibSp는 같이 승선한 형제자매와 사촌의 수를 나타내거나 같이 승선한 배우자수를 나타낸다. Parch는 같이 승선한 부모의 수를 나타낸다. 이 SibSp와 Parch 두 항목은 동행 가족없이 혼자 승선했는지를 판단하는 Alone이라는 항목을 생성하는데 사용했다. Ticket는 티켓번호를 의미하는데 사용하지 않았으며, Fare는 요금을 의미하는데 1-4까지의 4구간으로 변환했다. Cabin은 선실을 의미하는데 사용하지 않았다. Embarked는 출발항구를 의미하는데 Q, S, C로 되어있는데 이를 각각 0, 1, 2로 변환했다. 이와 같이 모두 수치로 변환해서 사용했다.

4. 모델의 구현 및 평가

4.1 모델구현

합성곱신경망모델은 [표 2]에 나타낸 바와 같이 입력층, 출력층외에 컨볼루션층 2개, 서브샘플링층 2개, 은닉층 2개로 총 8개 층으로 구현한다. 재난데이터는 1차원 자료여서 합성곱에서는 Conv1D, 서브샘플링에서는 MaxPooling1D를 사용한다. 그리고 활성함수로 ReLU를 사용하였으며, 출력층에서는 Softmax를 사용하였다.

[표 2] 합성곱신경망의 층구성
[Table 2] Layer Configuration of Convolutional Neural Networks

층	크기	활성함수
입력층	N=7*1	-
컨볼루션층1,2	2*1 필터 32개, 2*1 필터 32개	ReLU
서브샘플링층1,2	2 폴링필터	
은닉층	8 노드	ReLU
출력층	1노드	Softmax

모델에서 훈련된 각 층별 파라미터수를 별로 [표 3]에 나타냈다.

[표 3] 훈련된 파라미터 수
[Table 3] Number of Parameters Trained

층	파라미터 갯수 산출식	파라미터 갯수
C1	(필터사이즈*입력맵개수 + 바이어스)*특성맵개수	(2*1*1+1)*32 = 96
C3	(필터사이즈*입력맵개수 + 바이어스)*특성맵개수	(2*1*32+1)*32 = 2080
F6	(입력개수 + 바이어스)*출력개수	(32 +1)*8 = 264
F7	(입력개수 + 바이어스)*출력개수	(8+1)*1=9
누계		2449

4.2 평가결과

LeNet-5로 구현한 모델을 훈련데이터셋으로 훈련을 하고 테스트데이터셋으로 평가를 해야 하지만 Kaggle에서는 테스트데이터셋에 Survival항목을 포함시키지 않고 있다. 이 이유는 테스트데이터셋을 이용한 평가는 Kaggle 사이트에서 내부적으로 하기 위해서이다. 그리하여 어쩔 수 없이 최종 훈련결과를 바탕으로 훈련데이터셋를 이용하여 평가결과를 계산하였으며 [표 4]에 나타낸 바와 같이 나타났다. 훈련데이터셋에서 정확성은 85.63%, 정밀도는 92.35%를 기록하였다. 일반적으로 분류에서 사용되는 정확성과 정밀도 등의 평가척도를 계산하여 [표 5]에 나타냈다. 총 15층으로 구성한 Mamba의 분석 [9]에서는 정확성은 84.18%, 정밀도는 87.25%로 나타난 바 본 연구에서 구성한 LeNet-5가 더 효율적인 것을 알 수 있다. 훈련된 파라미터수를 비교하면 LeNet-5를 이용하여 본 연구에서 구성한 모델에서는 총 2,449개로 Mamba 분석에서의 총 87,617개의 약 2.8% 수준이어서 연산의 측면에서도 비교될 수 있다. 이는 일반적인 합성곱신경망에 대해서 LeNet-5가 가지는 장점이기도 하지만 그 수가 아주 작아서 연산에서도 월등히 효율적인 것을 알 수 있다.

[표 4] 평가결과

[Table 4] Results of Evaluation

구분		실제		
		사망	생존	소계
예측	사망	507	42	549
	생존	86	256	342
	소계	593	298	891

[표 5] 평가척도 비교

[Table 5] Comparisions of Performance Measure

평가척도	정의	본 연구	Mamba 모델
정확도(Accuracy)	(TP+TN)/(TP+FP+FN+TN)	0.8563	0.8418
정밀도(Precision)	TP/(TP+FP)	0.9235	0.8725
민감도(Sensitivity, Recall, True Positive Rate)	TP/(TP+FN)	0.8550	0.8709
특이도(Specificity, True Negative rate)	TN/(FP+TN)	0.8591	0.7947
False Positive Rate	FP/(FP+TN)	0.1409	0.2053
오류율(Error Rate)	(FP+FN)/(TP+FP+FN+TN)	0.1437	0.1582

TP: True Positive, FP: False Positive (유형 1오류), FN: False Negative(유형 2 오류), TN: True Negative

5. 결론

그 동안 타이타닉재난데이터에 대하여 k-최근접이웃 알고리즘, Naive Bayes 알고리즘, 의사결정트리 알고리즘, 앙상블 알고리즘(랜덤포리스트, xgboost 등), SVM 알고리즘, 로지스틱 회귀알고리즘, 신경망 알고리즘등의 인공지능기법을 이용한 많은 분석이 이루어져왔었다. 특히 신경망을 이용한 분석도 이루어졌지만 합성곱신경망을 이용한 분석을 없었다.

본 연구에서는 이 타이타닉재난데이터에 대하여 합성곱신경망중 LeNet-5를 이용하여 분석하였다. 이 합성곱신경망을 이용함으로써 그 동안 사용된 일반적인 다중계층신경망보다 학습파라미터수를 줄임으로서 학습의 효율을 기할 수 있었다. 그리고 이 결과에서 나타난 정확성, 정밀도 등의 평가척도에서 더 우수한 성능을 나타냈다.

향후 연구에서는 다른 형태의 신경망을 적용해서 분석해서 더 비교분석할 필요가 있다.

References

[1] wikipedia, “Titanic”, wikipedia.org, https://en.wikipedia.org/wiki/Titanic, (accessed November 2, 2020).

[2] W. H. Garzke, D. K. Brown, A. D. Sandiford, “The structural failure of the Titanic”, OCEANS '94: ‘Oceans Engineering for Today's Technology and Tomorrow's Preservation’, September 13-16, 1994, Brest, France, pp. 138-148, doi: 10.1109/OCEANS.1994.364186.

[3] D. A. Bright, R. M. L. Williams, A. S. McLaren, “Comparative photometric analysis of structural degradation on the bow of RMS Titanic”, OCEANS 2005 MTS/IEEE, September 17-23, 2005, Washington, DC, USA, pp. 106-110, doi: 10.1109/OCEANS.2005.1639746.

[4] P. K. Matthias, R. F. Silloway, “Use of automated seabed photomosaicing in forensic analysis of the RMS TITANIC disaster”, OCEANS 2000 MTS/IEEE Conference and Exhibition, September 11-14, 2000, Providence, RI, USA, pp. 667-671, doi: 10.1109/OCEANS.2000.881330.

[5] D. G. Kim, Y. S. Park, L. J. Park, T. Y. Chung, “Developing of New a Tensorflow Tutorial Model on Machine Learning : Focusing on the Kaggle Titanic Dataset”, IEMEK Journal of Embedded Systems and Applications, vol. 14, no. 3, August 2019, pp. 207-218, doi: 10.14372/IEMEK.2019.14.4.207.

[6] K. Singh, R. Nagpal, R. Sehgal, “Exploratory Data Analysis and Machine Learning on Titanic Disaster Dataset”, 2020 10th International Conference on Cloud Computing, Data Science & Engineering, January 29-31, 2020, Noida, India, pp. 320-326, doi: 10.1109/Confluence47617.2020.9057955.

[7] J. Shetty, S. Pallavi, Ramyashree, “Predicting the Survival Rate of Titanic Disaster Using Machine Learning Approaches”, 2018 4th International Conference for Convergence in Technology, October 27-28, 2018, Mangalore, India, pp. 1-5, doi: 10.1109/I2CT42659.2018.9058280.

[8] A. Singh, S. Saraswat, N. Faujdar, “Analyzing Titanic disaster using machine learning algorithms”, 2017 International Conference on Computing, Communication and Automation, May 5-6, 2017, Greater Noida,

India, pp. 406-411, doi: 10.1109/CCAA.2017.8229835.

[9] Black Bamba, "Getting Started with Titanic using neural network", https://www.kaggle.com/theblackmamba31/titanic-using-neural-network, (accessed November 10, 2020).

[10] Y. Lecun, L. Bottou, Y. Bengio, P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, vol. 86, no. 11, November 1998, pp. 2278-2324, doi: 10.1109/5.726791.

[11] Keras, "keras", keras.io, https://keras.io/, (accessed November 8, 2020).

[12] wikipedia, "k-nearest neighbors algorithm", wikipedia.org, https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm, (accessed November 12, 2020)

[13] wikipedia, "Naive Bayes classifier", wikipedia.org, https://en.wikipedia.org/wiki/Naive_Bayes_classifier, (accessed November 12, 2020)

[14] wikipedia, "Decision tree", wikipedia.org, https://en.wikipedia.org/wiki/Decision_tree, (accessed November 13, 2020)

[15] wikipedia, "Ensemble", wikipedia.org, https://en.wikipedia.org/wiki/Ensemble, (accessed November 13, 2020)

[16] wikipedia, "SVM", wikipedia.org, https://en.wikipedia.org/wiki/SVM, (accessed November 14, 2020)

[17] wikipedia, "Logistic regression", wikipedia.org, https://en.wikipedia.org/wiki/Logistic_regression, (accessed November 14, 2020)

[18] wikipedia, "Artificial neural network", wikipedia.org, https://en.wikipedia.org/wiki/Artificial_neural_network, (accessed November 14, 2020).

[19] Kaggle Inc., "kaggle", kaggle.com, www.kaggle.com, (accessed November 15, 2020).