ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 13 (2023) No. 2, pp. 501-514
DOI:10.33168/JSMS.2023.0234

Development of Online Fraud Detection and Sales Prediction Model using Supply Chain Dataset

Bong-hyun Kim

School of Software, Major of Computer Engineering, Seowon University, 377-3 Musimseo-ro, Seowon-gu, Cheongju-si, Chungcheongbuk-do, 28674, Korea,

bhkim@seowon.ac.kr

Abstract. With the global spread of smart devices, an environment that can be connected to the Internet anytime, anywhere is being established. The Internet of Things (IoT) is creating an environment in which not only people but also objects can be connected to networks. In addition, the upcoming future is being explained in various terms, ranging from the digital economy, the sharing economy, the on-demand economy, and the 4th industrial revolution. However, these changes show a common point in that they are all based on ICT technology. ICT is attracting attention in terms of promoting innovation in traditional industries, regardless of service and manufacturing, and presenting the possibility of creating new industries by enabling new business models. However, this new technology dissemination and activation part is being used in unexpected fields. Due to this, many damages such as fraud, phishing, and crime are caused online. In this paper, author designed and developed a deep learning model that detects sales fraud occurring online and predicts sales volume. To this end, a model was developed that receives the smart supply chain dataset provided by DataCo Global as an input and detects regions where sales fraud is detected, payment methods, and customers.

Keywords: Fraud Detection, Sales Prediction, Supply Chain Dataset, Tensorflow, Keras, XGBoost, LightGBM

1. Introduction

As the Internet of Things (IoT) has emerged as an emerging technology in recent years, more and more companies are using IoT technology. In addition, many IoT sensors are used to share real-time information, and a large amount of data is being generated (Park et al., 2021; Alyouzbaky et al., 2021). In particular, the evolution of ICT technology, such as the mobile Internet, connected devices including smartphones, and numerous Internet platform services, is changing many parts of the economy and society. As individuals and things are connected to the network, transactional activities such as product purchases are being actively conducted. In addition, all entertainment services such as broadcasting, publishing, movies, and music are also included in this phenomenon (Nam et al., 2022). Of course, production and consumption activities that are different from before are taking place, such as various services such as education, medical care, and finance through the Internet platform.

The upcoming future is being described in various terms, ranging from the digital economy, the sharing economy, the on-demand economy, and the 4th industrial revolution. However, these changes show a common point in that they are all based on ICT technology. ICT is promoting innovation in traditional industries, regardless of service and manufacturing. In addition, it is attracting attention in terms of suggesting the possibility of creating a new industry by enabling a new business model (Yang et al., 2022).

As ICT technology evolves, the paradigm in the service industry is showing the following paradigm shifts. First, the entry of ICT companies into the service industry is broadly expanding. As the importance of Internet platforms is emphasized due to the expansion of smartphone penetration, ICT companies that provide platforms are making inroads into the service industry. Device-based platforms such as Samsung and Apple, software-based platforms such as Google and Microsoft, as well as powerful platforms such as O2O and SNS are building an ecosystem that grows together through mutual connection. As the entry of ICT companies into heterogeneous industries has spread to a level that threatens traditional businesses, it is accelerating the collapse of the boundaries between industries. As a result, existing business operators and ICT companies, new entrants, are responding to the new environment while forming a competitive structure, while simultaneously collaborating with each other, such as voluntary collaboration, to fill each other's insufficient capabilities and acquire new know-how (Chen et al., 2020; Moon 2021).

Second, the spread of O2O services due to the collapse of on/offline boundaries. The reason for the rapid expansion of the O2O market is that customized information services have been provided so that consumers can easily access differentiated needs online due to the spread of smartphones and the development of ICT technologies such as the Internet of Things and location-based technologies. Currently, O2O is widely spreading in the form of on-demand in various service fields such as space, talent, and delivery as well as lodging, vehicles, and finance (Huangfu et al., 2020; Li 2020).

Third, innovative AI services based on data analysis are rapidly spreading. In recent years, it has been possible to secure large-scale real-time data, and with the rapid improvement of computing power function, etc., artificial intelligence services based on data analysis are on the rise. Recognizing the growth potential of artificial intelligence services, large ICT companies secure data from users through their platforms and provide innovative services by applying artificial intelligence technologies such as machine learning. Accordingly, each company is actively investing in long-term technology development investment, M&A promotion, and professional manpower recruitment to preoccupy the ecosystem. Recently, startups are also providing innovative artificial intelligence services. If certain patterns are analyzed by properly using a lot of data generated by the expansion of these new ICT technologies, it will be of great help to companies and companies in making better decisions in the future (Min et al., 2020).

Therefore, in this paper, we designed and developed a deep learning model for detecting online sales fraud and predicting sales volume using the smart supply chain dataset. To this end, a model was developed that receives the smart supply chain dataset provided by DataCo Global as an input and detects regions where sales fraud is detected, payment methods, and customers. In detail, customer segmentation analysis was performed using the smart supply chain dataset. In addition, we designed a deep learning model to detect online sales fraud and predict sales by using a deep learning framework and ensemble techniques.

2. Literature Review

2.1. Deep Learning Framework

A framework can be defined in the dictionary as "providing a series of classes in a collaborative form so that the design and implementation of specific parts of software can be reused." To put it simply, in programming, it can be seen as a basic framework and environment for creating a program (Son et al., 2020; Farhoudi et al., 2021). Figure 1 is a statistical graph of the preference for the framework used for deep learning development on the "towards data science" site.

A framework is a kind of package that bundles various libraries or modules for application development into one for efficient use. The Deep Learning Framework provides a number of libraries that have already been verified and various deep learning algorithms that have been pre-trained so that developers can use them quickly and easily. This frees developers from the exhausting work of implementing redundant features and helps them focus on developing core algorithms to solve problems (Wang et al., 2021).

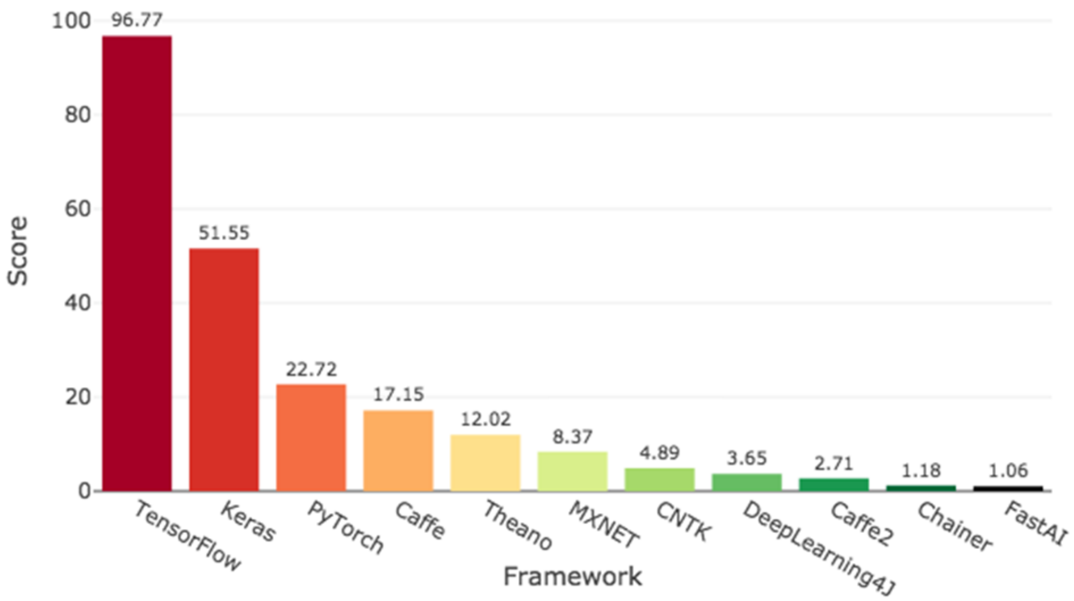

Fig. 1: Deep Learning Framework Power Scores

Deep learning framework refers to software that helps to develop deep learning models easily and quickly. It is also called by various names such as deep learning software, deep learning library, machine learning library, deep learning tool, and open source AI tool. Deep learning frameworks simplify the implementation of deep neural network (DNN) models, making it easy to create and test various deep architectures. Representative features of this framework include creating large and deep DNN computational models (model building), computing gradients within computational models (automatic differentiation), and packaging cuDNN, cuBLAS, etc. to run efficiently on GPUs. There are things that make it possible (using GPU), etc. Representative examples of deep learning frameworks shared as open source software include MXNet (Amazon), TensorFlow (Google), CNTK (CNTK, Microsoft), PaddlePaddle (Baidu), Caffe2 (Facebook), PyTorch, Facebook, etc (Kim et al., 2021; Yang et al., 2021).

2.1.1 TensorFlow

TensorFlow is an open source release by Google, built on years of experience learning from creating and using a previous model called DistBelief. It is designed to be flexible, efficient, scalable, and portable, so it can run on computers of any shape or size, from smartphones to computer clusters. Lightweight software is provided that allows trained models to be used in production in no time, and there is no need to create new models. TensorFlow adopts the innovation and collaboration of open source, but also has the support, help, and stability of big companies, so it can be used by individuals and enterprises, from startups to large companies like Google.

TensorFlow uses the Data Flow Graph method. The characteristic of this data flow graph is to express mathematical calculations and data flow as a directed graph using nodes and edges. Nodes perform tasks such as mathematical calculations, data input/output, reading, and writing, and edges represent relationships between nodes. Multi-dimensional data arrays of dynamic size are carried using the edge. This array is called a Tensor and represents the flow of Tensors (Wang et al., 2021).

Machine learning is a complex field, but the process of implementing machine learning models has been simplified by machine learning frameworks such as Google TensorFlow, which make it easy to collect data, train the model, make predictions, and then adjust the results. TensorFlow applications can run on virtually any target, including local machines, clusters in the cloud, iOS and Android devices, CPUs and GPUs. If you use Google Cloud, you can get even faster speeds by running TensorFlow on Google's custom TensorFlow Processing Unit (TPU) silicon. The resulting model created by TensorFlow can be deployed on most devices to handle prediction tasks.

2.1.2 Keras

Keras is a deep learning library used in representative Python along with Tensorflow. Keras is a library for Python built on top of machine learning & deep learning engines such as Tensorflow. Currently Keras belongs to the TensorFlow project, so the basic engine is recognized as TensorFlow, but in the past it was a library for Theano (Khow et al., 2022; Madhavi et al., 2022).

Keras is an open-source neural network library written in Python. It can be done on top of MXNet, Deeplearning4j, TensorFlow, Microsoft Cognitive Toolkit or Theano. Keras is a library that makes it easy to handle deep learning engines such as TensorFlow, MXNet, and Theano. In other words, it can be called a deep learning interface library made for more people to use deep learning. Figure 2 shows the structure of Keras. As you can see from the Keras architecture diagram, there are CPU, GPU, and TPU below TensorFlow, and Keras above. As such, if TensorFlow actually controls hardware, Keras is a concept that controls TensorFlow that controls hardware.

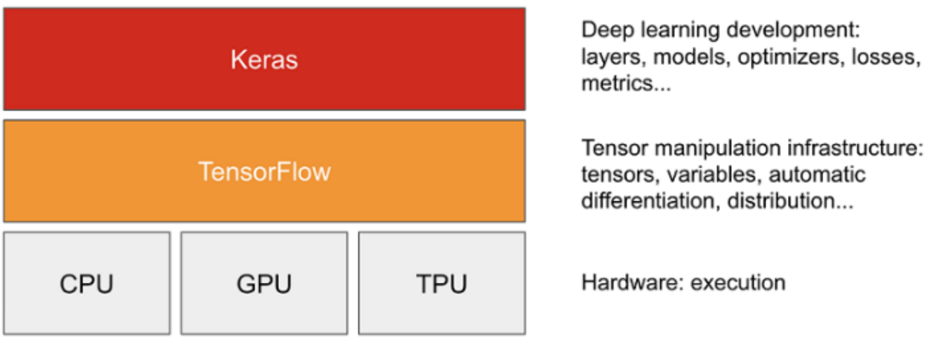

Fig. 2: Keras Structure

2.2. Ensemble Technique

Classification through ensemble learning is called a technique for deriving a more accurate final prediction by generating multiple classifiers and combining predictions. The goal of this learning is to combine the prediction results of various classifiers to obtain more reliable predictions than when using a single classifier.

For unstructured data such as images, videos, and voices, deep learning shows excellent performance when classifying, but for most structured data, ensemble techniques show excellent performance when classifying. Very famous methods in ensemble techniques include techniques such as "XGBoost", "LightGBM", and "Stacking" (Wang et al., 2020).

The types of ensemble learning are traditionally divided into voting, bagging, and boosting, and there are various other methods including stacking. What voting and bagging have in common is that multiple classifiers decide the final prediction result through voting. Voting is usually combining classifiers with different algorithms. In bagging, each classifier is the same type of algorithm. However, it is said that voting is performed by learning while taking data sampling differently. The most representative method of bagging is the Random Forest algorithm (a method using multiple decision trees). Bagging is used by sampling and extracting the training data assigned to each classifier from the original training data, as shown in the picture above, and this is called Bootstrapping. However, these sampled data are repeatedly sampled.

Boosting learns several classifiers sequentially, but proceeds with learning and prediction by assigning weights to the next classifier so that it can correctly predict data that is incorrectly predicted by the previously learned classifier. It has excellent predictive performance, leading to ensemble learning. Examples include Gradient boost, XGBoost, and LightGBM. Stacking is a method of predicting results by making prediction results of various different models into training data again and retraining with other models (meta-models). Due to the method of stacking results, it is marked as stacking (Zhang et al., 2019; Sahoo et al., 2019).

2.2.1 XGBoost

XGBoost is an ensemble algorithm that uses a combination of several decision trees. eXtreme Gradient Boosting (XGBoost) is an open-source implementation of the Gradient Boosted Decision Trees algorithm. XGBoost provides models such as classification and regression (Lahoti et al., 2018).

The XGBoost library also has its own API, but uses the scikit-learn wrapper classes XGBRegressor and XGBClassifier. To prepare the data and evaluate the model using the scikit-learn library. XGBRegressor is an XGBoost model for regression, and XGBClassifier is an XGBoost model for classification. The characteristics of XGBoost are as follows.

- XGBoost is faster than Gradient Boosting Analysis (GBM).
- Regulations are included to prevent overfitting.
- Based on CART (Classifier and Regression Tree). (i.e. both regression and classification are possible)
- Provide early stopping.
- Weighting, which is a feature of boosting, is done by gradient descent.

2.2.2 LightGBM

LightGBM (Light Gradient Boosting Machine) is a relatively recent algorithm, and it is made of a tree structure based on gradient boosting. While other algorithm trees expand horizontally, lightGBM has a structure in which the tree expands vertically. LightGBM is a method of predicting the result value by using several different models instead of using one powerful model (Chen et al., 2019).

The main idea of LightGBM is twofold. One is a method called GOSS (Gradient-based One-Side Sampling), and the other is a method called EFB (Exclusive Feature Bundling). Both methods have in common that they are 'how to effectively reduce the size of data without affecting the distribution of given data' (Weng et al., 2019). The features of LightGBM are as follows.

- LightGBM learns faster than XGBoost, and the prediction execution time is also fast.
- Has smaller memory usage.
- Automatic conversion and optimal segmentation of categorical features.

(It converts categorical variables optimally without using one-hot encoding and performs node division accordingly.)

- When applied to a small data set, overfitting is likely to occur.

3. Data Collection and Analysis

3.1. Data Collection

In this paper, a study was conducted to derive regions, payment methods, and customers where sales fraud was detected using the supply chain data set used by DataCo Global. For data collection, import the Pandas library for Python data processing and the Numpy package for handling numerical data and import Matplotlib and Seaborn to visualize data. Import ensemble algorithm XGBoost, LightGBM. Through ensemble learning, a string classifier (strong classifier) can be created by combining several weak classifiers. Also, import the datetime module to handle date and time data. Finally, import TensorFlow, which provides various functions to easily implement deep learning.

The data set consists of about 180,000 data from the supply chain used by DataCo Global for 3 years and is provided and used by DataCo. For data loading, use pandas read csv method to load csv data. Figure 3 is a screen displaying only the top 10 collected data.

	Type	Days for shipping (real)	Days for shipment (scheduled)	Benefit per order	Sales per customer	Delivery Status	Late_delivery_risk	Category Id	Category Name	Customer City	...	Order Zipcode	Product Card Id	Product Category Id	D
0	DEBIT	3	4	91.250000	314.640015	Advance shipping	0	73	Sporting Goods	Caguas	...	NaN	1360	73	
1	TRANSFER	5	4	-249.089996	311.359985	Late delivery	1	73	Sporting Goods	Caguas	...	NaN	1360	73	
2	CASH	4	4	-247.779999	309.720001	Shipping on time	0	73	Sporting Goods	San Jose	...	NaN	1360	73	
3	DEBIT	3	4	22.860001	304.809998	Advance shipping	0	73	Sporting Goods	Los Angeles	...	NaN	1360	73	
4	PAYMENT	2	4	134.210007	298.250000	Advance shipping	0	73	Sporting Goods	Caguas	...	NaN	1360	73	
5	TRANSFER	6	4	18.580000	294.980011	Shipping canceled	0	73	Sporting Goods	Tonawanda	...	NaN	1360	73	
6	DEBIT	2	1	95.180000	288.420013	Late delivery	1	73	Sporting Goods	Caguas	...	NaN	1360	73	
7	TRANSFER	2	1	68.430000	285.140015	Late delivery	1	73	Sporting Goods	Miami	...	NaN	1360	73	
8	CASH	3	2	133.720001	278.589996	Late delivery	1	73	Sporting Goods	Caguas	...	NaN	1360	73	
9	CASH	2	1	132.149994	275.309998	Late delivery	1	73	Sporting Goods	San Ramon	...	NaN	1360	73	

Fig. 3: Collected Dataset

3.2. Data Analysis

Data cleaning was performed prior to data analysis. In particular, create a new column containing the full customer name to avoid name duplication. Based on the refined data, visualization for data analysis was performed. A heatmap was derived using the fig and ax variables used in the visualization process. Figure 4 shows the visualization of the data as a heatmap.

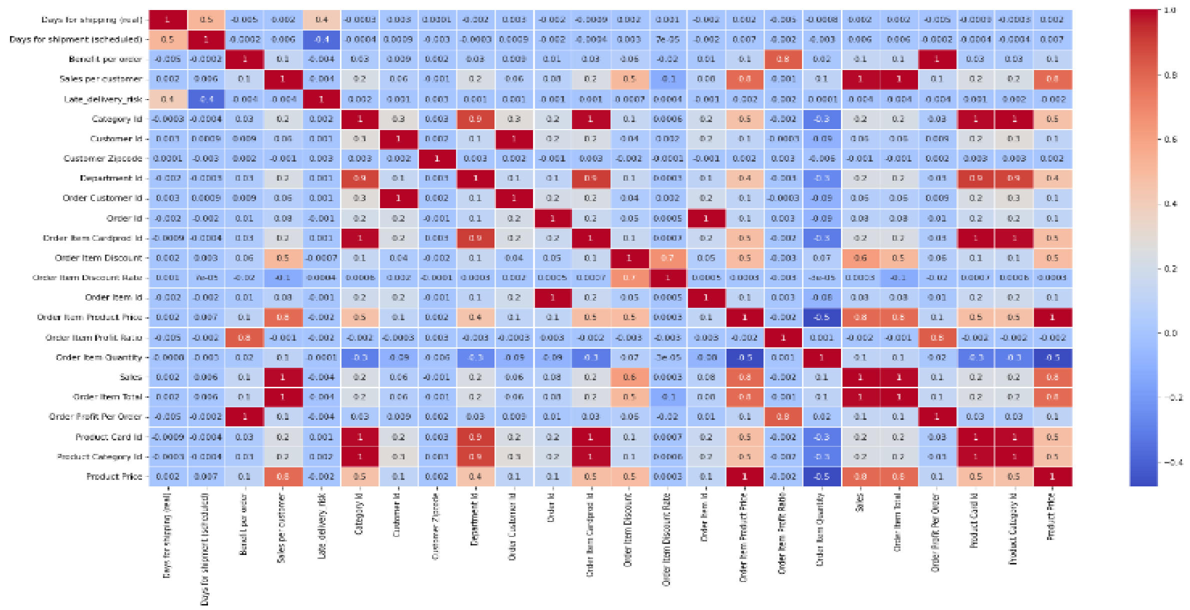

Fig. 4: Data Visualization using Heatmap

Based on the results of data visualization through heatmap, it can be seen that products and prices show a high correlation with sales and the total number of ordered items. In addition, markets and regions were grouped to check all sales in the region in a bar-shaped graph by combining similar markets and regions. Figure 5 and 6 show all sales by region in the form of a bar graph using plot.bar.

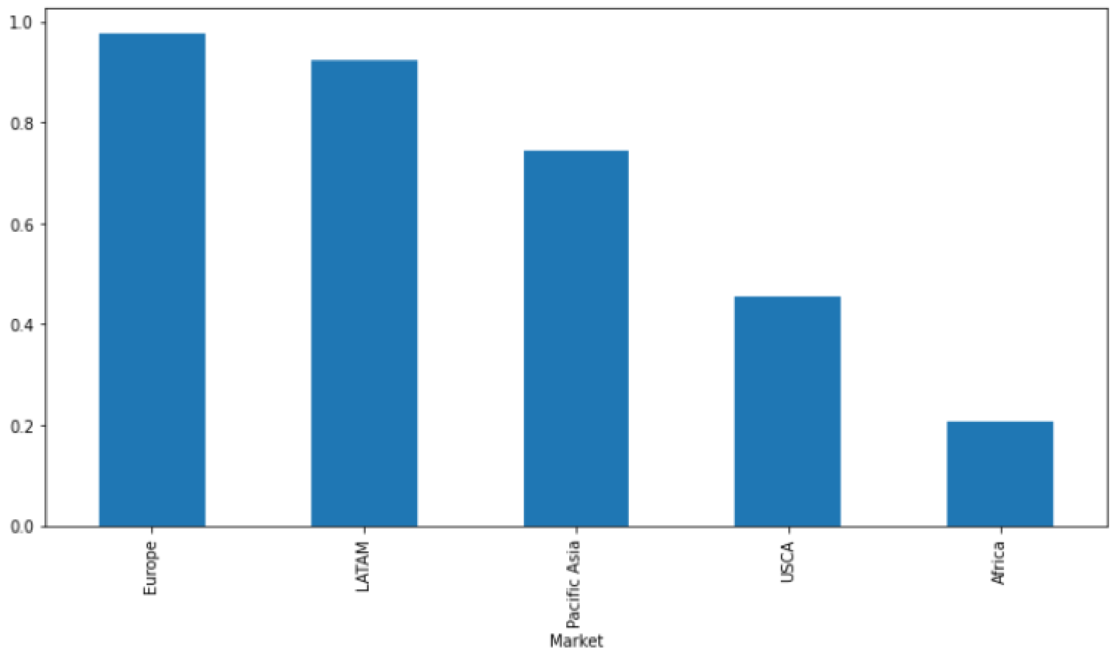

Fig. 5: Total Sales for All Markets

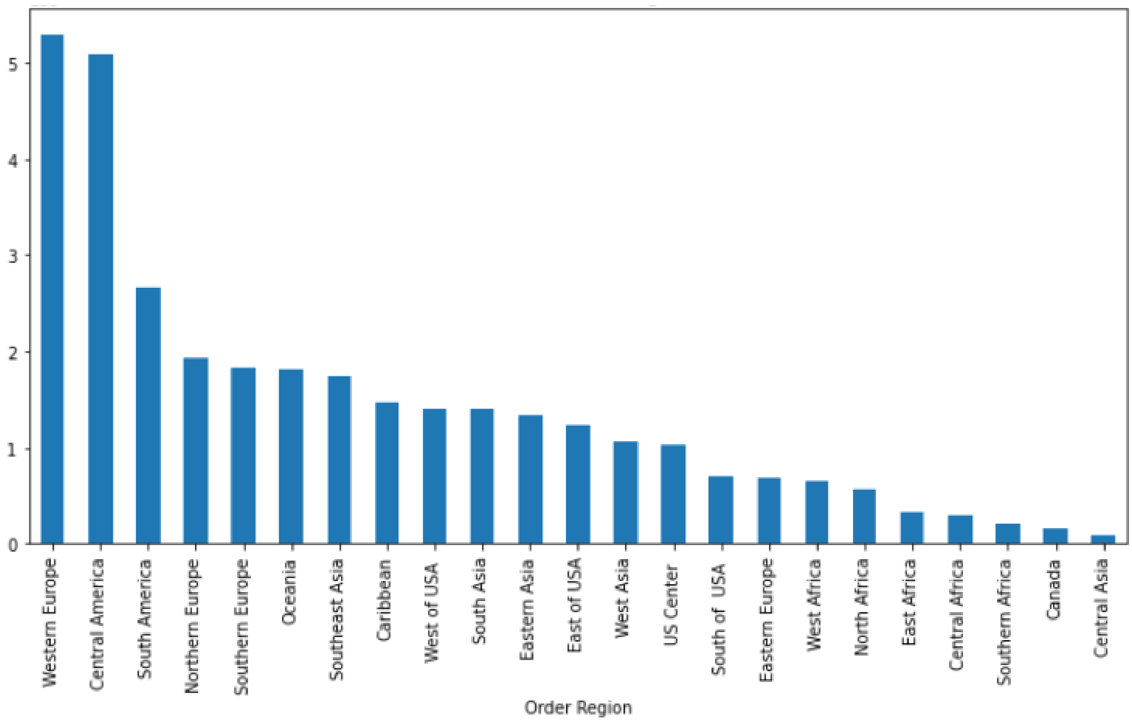

Fig. 6: Total Sales for All Regions

Looking at the graph of the results by region for total sales, it can be seen that the European market has the most. Also, Africa had the lowest sales, while Western Europe and Central America had the highest sales. In the same way, product category analysis was performed. Figure 7 breaks down total sales for products sold.

As can be seen from the analysis results, the fishing category is recording the highest sales. The top seven products with the highest average price are the best-selling products. Because there is a strong correlation between price and sales, it is important to understand how the price of a product affects sales. Figure 8 shows the correlation analysis between product price and sales.

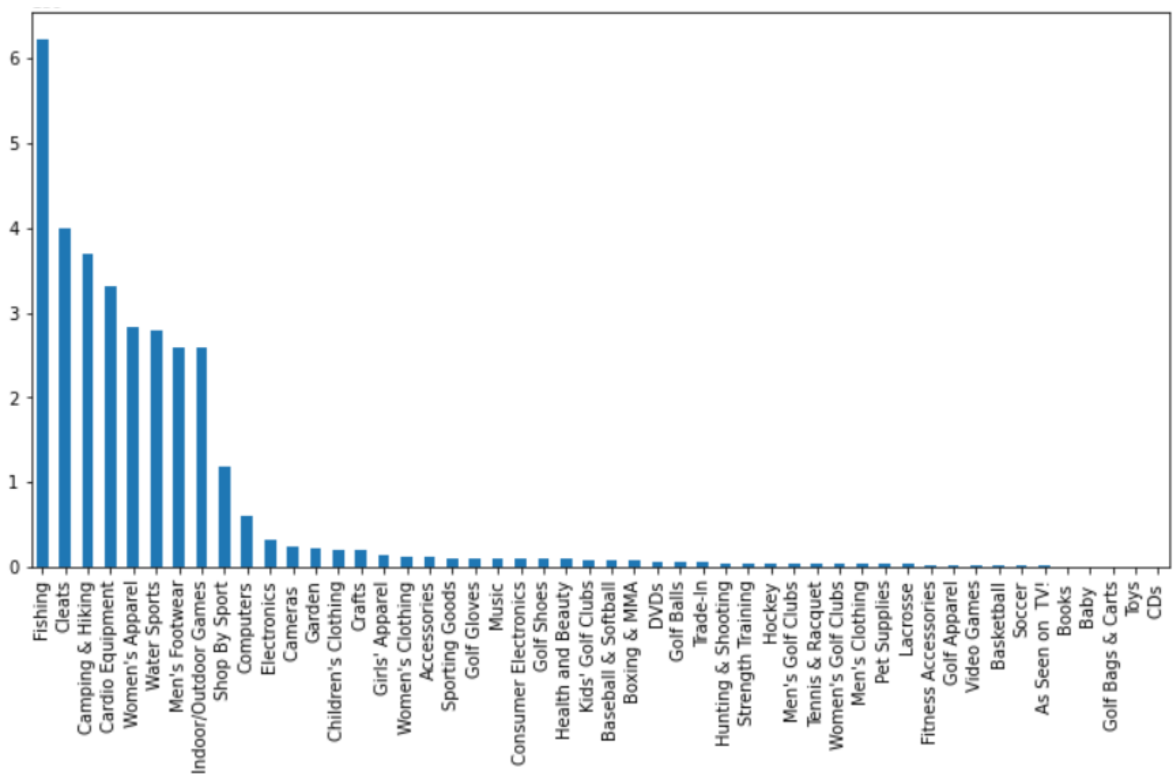

Fig. 7: Total Sales by Product Category

As a result of the analysis, it can be confirmed that product price has a linear relationship with sales.

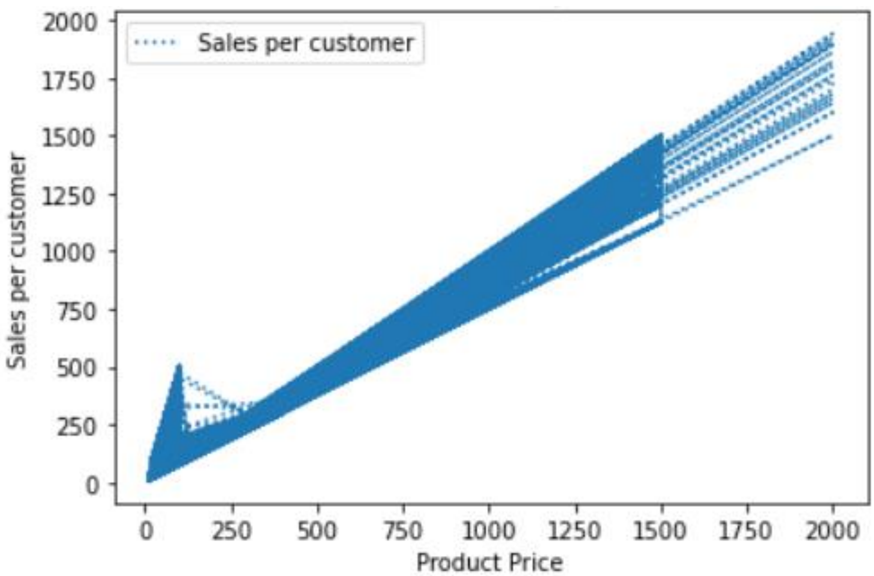

Fig. 8: Correlation Analysis between Product Price and Sales

In addition, the sales status by period was also analyzed. Figure 9 breaks down average sales by year, week, hour, and month. As a result of the analysis, the month with the most orders was October, followed by November. Orders for all other months are constant. It can be seen that the year with the highest number of orders is 2017, and Saturday has the highest average sales volume and Wednesday has the lowest.

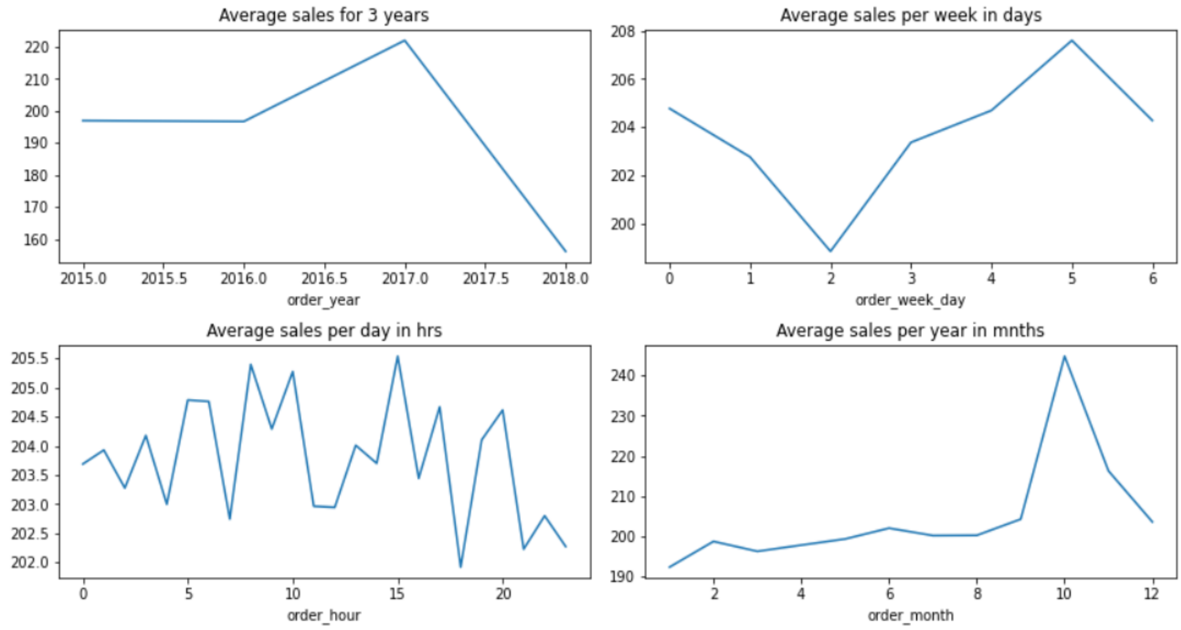

Fig. 9: Analysis of Sales Status by Period

Knowing the payment method used when committing sales fraud is useful in preventing fraud. Therefore, the unique() method was used to analyze the 4 ways customers paid. As a result of the analysis, it can be seen that the most preferred payment method for people in all regions is direct payment, and cash payment is not preferred. Some products have negative returns, which causes the company to lose revenue.

Total lost sales are about 3.9 million, and the category with the most lost sales is Kletz. It is followed by men's shoes. Lost sales are mostly in Central America and Western Europe, and may be due to sales fraud or delivery delays. Since there are no fraudulent transactions in the DEBIT, CASH, and Payment methods, it can be seen that all fraudulent transactions are conducted using overseas remittances. Orders suspected of being fraudulent were isolated and regions with the most fraud were visualized in a pie chart. Figure 10 visualizes the percentage of orders that fall under sales fraud. As a result of the analysis, the most suspected sales fraud is Western Europe, accounting for 17.4% of the total.

4. Predictive Modeling and Discussion

In this paper, we measured the performance of various models for data prediction simulation. To this end, a machine learning model was trained to detect fraud and delayed delivery for the classification types. In addition, the regression model was applied to predict sales and order quantity. A copy of the data was created and used to train the data. Then, we converted the data to int type using the label encoder library so that it can be trained in the machine learning model.

Figure 11 is a simulation result of data modeling by applying random forest classification and logistic classification. In addition,

Figure 12 shows the simulation results applying the gradient boosting classification.

Finally, sales and order quantities were predicted through comparison of regression models. To do this, we use the MinMax scaler to standardize the data.

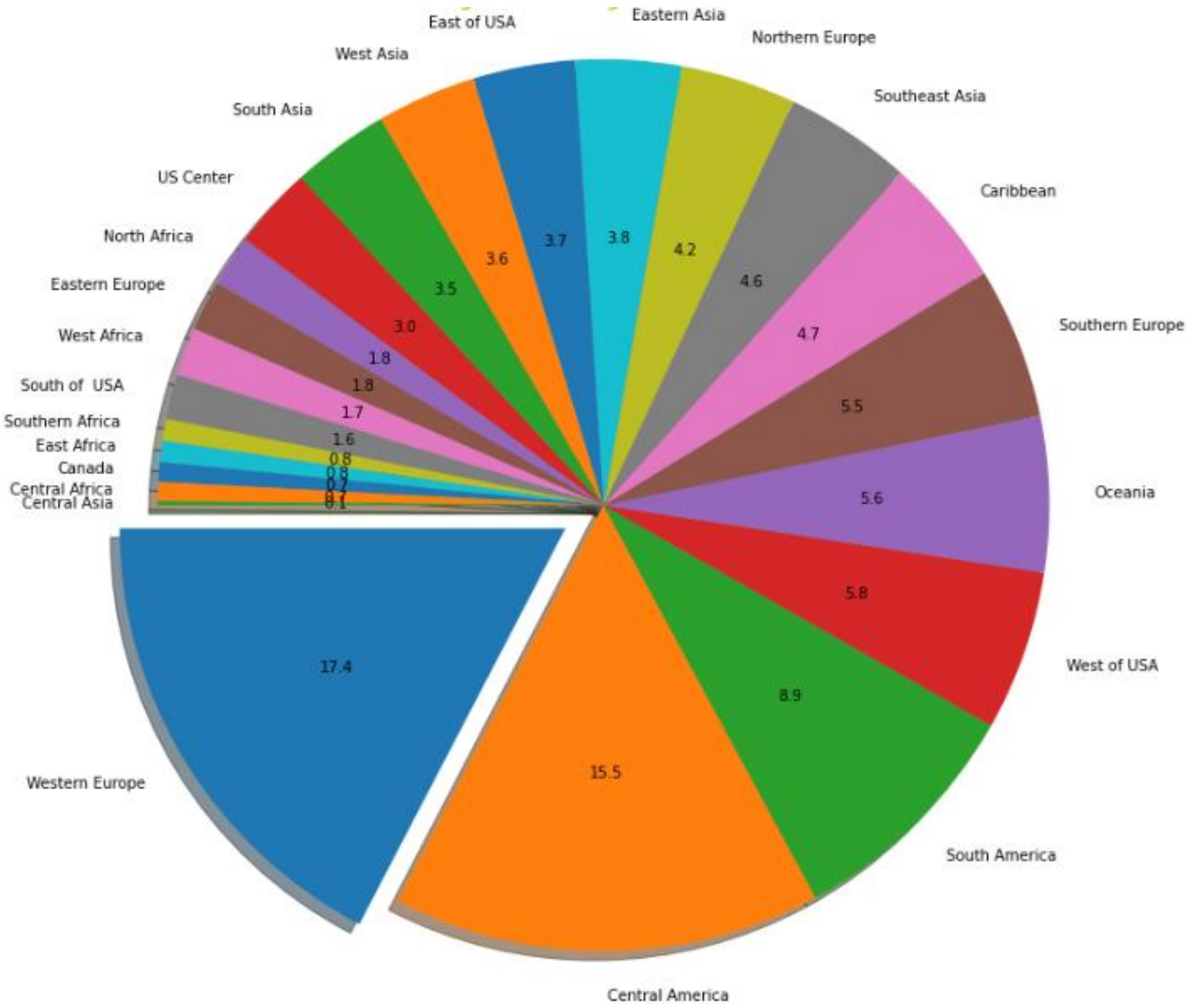

Fig. 10: Percentage of Sales Fraud Orders by Region

<Random Forest Classification>

```
Model paramters used are : RandomForestClassifier()
Accuracy of fraud status is          : 98.7065145136273 %
Recall score of fraud status is          : 98.72773536895674 %
Conf Matrix of fraud status is          :
 [[35249     5]
 [  462   388]]
F1 score of fraud status is          : 62.429605792437656 %
Accuracy of late delivery status is: 98.23842233547529 %
Recall score of late delivery status is: 96.88738804874468 %
Conf Matrix of late delivery status is:
 [[15671   636]
 [    0 19797]]
F1 score of late delivery status is: 98.41909023117077 %
```

<Logistic Classification>

```
Model paramters used are : LogisticRegression(random_state=0)
Accuracy of fraud status is          : 97.79248836693995 %
Recall score of fraud status is          : 58.307210031347964 %
Conf Matrix of fraud status is          :
 [[35121   133]
 [  664   186]]
F1 score of fraud status is          : 31.822070145423435 %
Accuracy of late delivery status is: 98.84777309993352 %
Recall score of late delivery status is: 97.9419185672587 %
Conf Matrix of late delivery status is:
 [[15891   416]
 [    0 19797]]
F1 score of late delivery status is: 98.96025993501625 %
```

Fig. 11: Data Modeling using Random Forest and Logistic

Figure 13 is the result of simulating data prediction by applying various techniques. As shown in Figure 13, the most improved performance was derived when the linear regression model was applied.

As a result, in case of predicting the order quantity, random forest and eXtreme Gradient Boosting showed good performance. In other words, it can be evaluated as a model with improved performance that can be used for order quantity prediction.

5. Conclusion and Future Work

```
Model paramters used are : XGBClassifier(base_score=0.5, booster='gbtree', callbacks=None,
              colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1,
              early_stopping_rounds=None, enable_categorical=False,
              eval_metric=None, feature_types=None, gamma=0, gpu_id=-1,
              grow_policy='depthwise', importance_type=None,
              interaction_constraints='', learning_rate=0.300000012,
              max_bin=256, max_cat_threshold=64, max_cat_to_onehot=4,
              max_delta_step=0, max_depth=6, max_leaves=0, min_child_weight=1,
              missing=nan, monotone_constraints='()', n_estimators=100,
              n_jobs=0, num_parallel_tree=1, predictor='auto', random_state=0, ...)
Accuracy of fraud status is        : 98.8671615333481 %
Recall score of fraud status is        : 90.31078610603291 %
Conf Matrix of fraud status is        :
 [[35201    53]
 [  356   494]]
F1 score of fraud status is        : 70.72297780959198 %
Accuracy of late delivery status is: 99.1746066917793 %
Recall score of late delivery status is: 98.52670349907919 %
Conf Matrix of late delivery status is:
```

Fig. 12: Data Modeling using Gradient Boosting

<Random Forest Regression>

```
Model parameter used are: RandomForestRegressor(max_depth=10, random_state=40)
MAE of sales is          : 0.19170857280830786
RMSE of sales is         : 1.882911557250712
MAE of order quantity    : 0.00011910028905403445
RMSE of order quantity   : 0.006147288482544812
```

<Light Gradient Boosting Regression>

```
Model parameter used are: LGBMRegressor()
MAE of sales is          : 0.4608967314236202
RMSE of sales is         : 1.6694322183572285
MAE of order quantity    : 0.0015026406737609238
RMSE of order quantity   : 0.01083017104892248
```

<eXtreme Gradient Boosting Regression>

```
Model parameter used are: XGBRegressor(base_score=0.5, booster='gbtree', callbacks=None,
             colsample_bylevel=1, colsample_bynode=1, colsample_bytree=1,
             early_stopping_rounds=None, enable_categorical=False,
             eval_metric=None, feature_types=None, gamma=0, gpu_id=-1,
             grow_policy='depthwise', importance_type=None,
             interaction_constraints='', learning_rate=0.300000012, max_bin=256,
             max_cat_threshold=64, max_cat_to_onehot=4, max_delta_step=0,
             max_depth=6, max_leaves=0, min_child_weight=1, missing=nan,
             monotone_constraints='()', n_estimators=100, n_jobs=0,
             num_parallel_tree=1, predictor='auto', random_state=0, ...)
MAE of sales is          : 0.15308655168496887
RMSE of sales is         : 3.132140470959337
MAE of order quantity    : 0.0006108814842882867
RMSE of order quantity   : 0.005077303885212369
```

<Linear Regression>

```
Model parameter used are: LinearRegression()
MAE of sales is          : 0.0005995640729386036
RMSE of sales is         : 0.0014783292416029614
MAE of order quantity    : 0.3420143704097005
RMSE of order quantity   : 0.5215045989826141
```

Fig. 13: Data Prediction Simulation

As COVID-19 has increased the amount of time spent at home, daily and social life has been affected accordingly. In addition to this, with the development of ICT convergence technology, all activities can be processed online. Under these circumstances, online shopping has become an important shopping means for people. Compared to traditional e-commerce, online commerce with both business and social functions is rapidly emerging as an important form of online shopping in modern society.

The outbreak of COVID-19 is having a great impact on consumer psychology and consumption environment. Consumers are increasingly experiencing feelings of loneliness, boredom, and anxiety in these social phenomena, and a phenomenon is occurring in which they are relieved in the form of online shopping. In this situation, cases of abusing the form of online shopping are frequently occurring. In particular, illegal activities such as sales fraud and information leakage continue to take advantage of the online nature. Therefore, research is needed to predict sales fraud and illegal activities in the online environment to solve this problem.

In this paper, fraud detection and sales prediction were performed using DataCo Global's data set. Analysis of the data showed that although sales were highest in Western Europe and Central America, they also suffered significant sales losses. Fraudulent transactions were suspected in both locations. The company's total sales continued through 2017, but we could see that sales declined in 2018. Most of the customers preferred to pay with a debit card, and we were able to derive a graph that showed many fraudulent transactions from one customer.

In addition, when predicting order quantity, random forest and eXtreme Gradient Boosting showed good performance, and it can be seen as a good model for order quantity prediction. Through the data modeling design and predictive simulation results in this paper, it is possible to derive a method to predict customer analysis and sales volume using deep learning.

References

Park S. H., Kim K. Y., Jung H. S., & Hyun D. W. (2021). A study on the overseas expansion strategy of domestic ICT companies in the post-corona era -Focusing on case studies of ICT companies. Journal of Digital Convergence, 19(10), 163-173. https://doi.org/10.14400/JDC.2021.19.10.163

Alyouzbaky, B. A., Hanna, R. D., & Najeeb, S. H. (2021). The effect of information overload, and social media fatigue on online consumers purchasing decisions: the mediating role of technostress and information anxiety. Journal of Logistics, Informatics and Service Science, 8(2), 19-46. https://doi.org/10.33168/LISS.2021.0202

Nam, J., & Choi, M. (2022). IoT edge cloud platform with Revocatable blockchain smart contract. Journal of Logistics, Informatics and Service Science, 9(2), 131-144. https://doi.org/10.33168/LISS.2022.0208

Yang, J., Liu, J., Guo, Z., Gao, L., Hadjikakou, M., & Bryan, B. (2022). A Multiple Nonlinear Regression Model for Estimating Input-Output Values with Partial Information. Economic Computation and Economic Cybernetics Studies And Research, 56(3), 169-186. https://doi.org/10.24818/18423264/56.3.22.11

Chen, X., & Hu, J. (2020). ICT-related behavioral factors mediate the relationship between adolescents' ICT interest and their ICT self-efficacy: Evidence from 30 countries. Computers and Education, 159, 104004. https://doi.org/10.1016/j.compedu.2020.104004

Moon S. H. (2021). A Study on ICT Conversion and Change of Industrial Society. The journal of the convergence on culture technology, 7(4), 653-658. https://doi.org/10.17703/JCCT.2021.7.4.653

Huangfu, S., Xue, X., & Wang, J. (2020). Resource-constrained O2O service recommended strategy research. International Journal of Services Technology and Management, 26(2), 202. https://doi.org/10.1504/ijstm.2020.106746

Li, L. (2020). Wechat Service of University Library under the Mode of O2O Reading Promotion. Advances in Journalism and Communication, 8(2), 49-58. https://doi.org/10.4236/ajc.2020.82004

Min S. H., & Son K. H. (2020). Comparative Analysis on ICT Supply Chain Security Standards and Framework. Journal of the Korea Institute of Information Security and Cryptology, 30(6), 1189-1206. https://doi.org/10.13089/JKIISC.2020.30.6.1189

Son, Y., Lee, Y.S. (2022). A smart contract weakness and security hole analyzer using virtual machine based dynamic monitor. Journal of Logistics, Informatics and Service Science, Vol. 9, No. 1, pp. 36-52. https://doi.org/10.33168/LISS.2022.0104

Farhoudi, Z., & Setayeshi, S. (2021). Fusion of deep learning features with mixture of brain emotional learning for audio-visual emotion recognition. Speech communication, 127, 92-103. https://doi.org/10.1016/j.specom.2020.12.001

Wang, P., Fan, E., & Wang, P. (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern recognition letters, 141, 61-67. https://doi.org/10.1016/j.patrec.2020.07.042

Kim Y. H., & Kim N. G. (2021). Deep Learning-based Pes Planus Classification Model Using Transfer Learning. Journal of Korean Computer Information Society, 26(4), 21-28. https://doi.org/10.9708/jksci.2021.26.04.021

Yang, X., Zhang, Y., Lv, W., & Wang, D. (2021). Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier. Renewable energy, 163, 386-397. https://doi.org/10.1016/j.renene.2020.08.125

Wang, J; Chen, Y; Qiu, S; Cui, Q. (2021). Cuckoo Search Optimized Integrated Framework Based on Feature Clustering and Deep Learning for Daily Stock Price Forecasting, Economic Computation and Economic Cybernetics Studies And Research, Vol. 55(3), pg. 55-70, DOI: 10.24818/18423264/55.3.21.04

Khow, Z.J., Goh, K.O.M., Tee, C., Law, C.Y. (2022). A YOVO5 based real-time helmet and mask detection system. Journal of Logistics, Informatics and Service Science, Vol. 9, No. 3, pp. 97-111. https://doi.org/10.33168/LISS.2022.0308

Madhavi, S., Hong, S.P. (2022). Anomaly detection using deep neural network quantum encoder. Journal of Logistics, Informatics and Service Science, Vol. 9, No. 2, pp. 118-130. https://doi.org/10.33168/LISS.2022.0207

Wang, C., Deng, C., & Wang, S. (2020). mbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. Pattern recognition letters, 136, 190-197. https://doi.org/10.1016/j.patrec.2020.05.035

Zhang, W., Zhao, X., & Li, Z. (2019). A Comprehensive Study of Smartphone-Based Indoor Activity Recognition via Xgboost. IEEE Access, 7, 80027-80042, https://doi.org/10.1109/access.2019.2922974

Sahoo, D., & Balabantaray, R. C. (2019). Single-Sentence Compression using XGBoost. International Journal of Information Retrieval Research, 9(3), 1-11. https://doi.org/10.4018/ijirr.2019070101

Lahoti, P., & Kumar, A. (2018). Imbalanced Data Classification using Sampling Techniques and XGBoost. International Journal of Computer Applications, 182(12), 19-22. https://doi.org/10.5120/ijca2018917735

Chen, C., Zhang, Q., Ma, Q., & Yu, B. (2019). LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. Chemometrics and intelligent laboratory systems : an international journal sponsored by the Chemometrics Society, 191, 54-64. https://doi.org/10.1016/j.chemolab.2019.06.003

Weng, T., Liu, W., & Xiao, J. (2019). Supply chain sales forecasting based on lightGBM and LSTM combination model. Industrial Management and Data Systems, 120(2), 265-279. https://doi.org/10.1108/imds-03-2019-0170