Analysing the Effects of Weather Conditions on Container Terminal Operations Using Machine Learning

Üstün Atak, Istanbul Technical University, Turkey*
https://orcid.org/0000-0002-1513-7371

Tolga Kaya, Istanbul Technical University, Turkey
Yasin Arslanoğlu, Istanbul Technical University, Turkey
https://orcid.org/0000-0002-9492-2975

ABSTRACT

Container ships transport a large amount of valuable cargo, and there is a demand for less expensive and faster transportation options. Weather, vessel type, and the nature and amount of the goods are all external elements that might impact container handling times, which are directly related to overall port stay time. In this scope, container terminal operations could be optimised with the help of historical data which provides access to classification and prediction of the cargo handling operations. In this study, the real-time data of a container terminal operation is analysed with different machine learning techniques along with the Fuzzy C-Means clustering method. The results show that Fuzzy C-Means clustering has a positive impact on the explanatory power of models in container terminal operations. The research revealed that an increase in wind speed influences cargo handling time for mobile cranes.

KEYWORDS

Container Terminal, Data Mining, Fuzzy Clustering, Machine Learning, Maritime Transportation

INTRODUCTION

Maritime transportation has been evolving at a rapid pace for decades, and the sector is directly impacted by global concerns. Safety, security, and environmental management have been at the forefront of these developments and difficulties. By imposing new laws or regulations, the maritime authorities aim to create a safety culture as well as clean waters for shipping. On the other hand, optimization solutions for stakeholders ensure safe, secure, and cost-effective maritime transportation choices.

One of the research areas of port management is the expected time of arrival time prediction (Dobrkovic et al., 2016; Meijer, 2017; Yu et al., 2018). Although voyage optimization has been a maritime topic for decades, port operations optimization is also a popular topic (Cammin et al., 2020). Vessels could transport cargo efficiently from one port to another if cargo handling time was minimized. Furthermore, optimized cargo operations and voyages result in a lower carbon footprint, more environmentally friendly transportation, and increased profits.
This research aims to analyse container port operations using historical data in this manner. The study dataset was gathered from several types of container vessels that conducted one or more visits to the sample container terminal in Turkey. Four mobile gantry cranes and two berthing quays are available at the sample container port. Because of its geographic position, the port’s operation is influenced by a variety of conditions, including humid, hot, and occasionally windy weather.

The objective of this study is to investigate several machine learning algorithms for analysing container handling time based on operational differences including crane properties, container cargo condition, container weight, operation types (loading, discharging, or shifting), and weather conditions. In summary, the study’s goals are as follows:

- Providing an overview of potential weather effects on container handling operations.
- Providing high-accuracy machine learning techniques that could be used to analyse container port operations with historical data.
- Exploring the key factors that could cause a delay in cargo operation time.
- Exploring possible solutions to reduce vessel port stay time.

Multiple Linear Regression (MLR), Ridge Regression, LASSO Regression, Principal Components Regression (PCR), Partial Least Squares Regression (PLS), Support Vector Regression (SVR), Random Forest Regression (RF), and Multilayer Perceptron Regression (MLP) methods are used in this paper to determine the effects of weather conditions. Furthermore, Fuzzy C-Means clustering (FCM) methods are used to analyse vessels’ port of calls in terms of operational benchmarks, such as cargo completion time, cargo handling capacity, and vessel dimensions.

The rest of the paper is organized as follows: previous studies in the literature are discussed in the second part. The methodology of the investigation is presented in Section 3. The container terminal case study is examined in the fourth part. The findings are evaluated, and the discussion is given in Section 5. Finally, conclusions are discussed.

**LITERATURE**

Fuzzy logic is a well-known topic in the maritime transportation industry, and it has a variety of applications ranging from accident investigation to decision support systems. The FCM clustering approach has been examined less frequently in the context of unsupervised classification techniques than fuzzy logic. There are just a few studies linked to the FCM algorithm in the marine area from 1999 to 2021. Sea-ice classification (Eom, 1999), typhoon track classification (Kim et al., 2011), detection of sea surface temperatures (Qin et al., 2010; Sousa et al., 2008; Tamim et al., 2019), and seafloor classification (Lucieer and Lucieer, 2009; Zare et al., 2018) are some of the studies that have used the FCM. Ship detection, fuel consumption, engine performance, collision avoidance, defect diagnosis, categorization of automatic radar plotting aid targets, and satellite images are also all being studied in the marine transportation field.

With the aid of the FCM clustering method, the investigations attempted to evaluate fundamental collision risk in practice and correct manoeuvre alternatives for master mariners (Ma et al., 2015). Similarly, Hu et al. (2020) employed Automatic Identification System (AIS) data to solve the danger of navigational collision. To achieve the clusters that are connected to the position and speed of fishing vessels, the researchers employed the FCM clustering approach. The study’s goal is to quantify fundamental collision risk and develop a collision risk algorithm for e-navigation. Similarly, Ma et al. (2015) employed FCM to build an intelligent radar target classification and identification system. The focus of the research is to figure out the likelihood of the target vessel’s speed over ground and position. The membership functions of each attribute are built using expert knowledge, statics, and electronic chart data. The authors’ enhanced FCM, in particular, is capable of reliably classifying radar objects.
Similarly, calculation and prediction of the main engine performance and fault diagnosis are studied by using FCM. The merchant vessels are navigating with the optimal speed to keep the balance between profit and costs due to the high price of fuel oil. To achieve the best optimal navigation speed, the main engine should run at high performance without any fault which causes delays. In this view, Chan and Chin (2017) studied the machine learning data analysis to model main engine performance by using FCM and K-means to reduce the root mean square error as well as methods such as neural network, MLR, and bagged regression. The purpose of the study is to predict a more robust and accurate model of the real-time engine. Another performance measure technique of the marine diesel engine is to calculate the optimal loading condition of the vessel. The study is conducted by Tran (2020) aims to achieve a multicriteria decision-making process that could be used for determining the optimal loading condition of the vessel with FCM. The author analysed wind speed, ship speed, wave height, and shaft speed to optimize the ship’s main engine load. To keep the propeller run at full performance, the fault diagnosis of the main engines is crucial in marine navigation. To predict fault conditions and take precautions in advance, the diagnosis of the marine diesel engine is the key element of the vessel’s voyages. To handle these challenges, Peng et al. (2012) proposed a method using FCM and grey relation analysis. Since the relationship between characteristic parameters and the reason of fault on the marine engine are complicated, the diagnosis is explained as grey, fuzzy, and nonlinear. In maritime transportation, the research studies are diverse from prediction to pattern identification in the view of machine learning techniques. Duca et. al (2017) studied the number of AIS messages to build new strategies for decision support systems. The study aimed to predict ship routes with K-Nearest Neighbor. The real data from AIS which was collected from the Malta vicinity analysed with the proposed algorithm. The results showed that the algorithm provides an accuracy of 0.931. Wen et. al (2014) proposed a software named RouteMiner which calculates ship routes from a set of trajectories. The authors indicated that the software provides a better understanding of the navigation data for maritime traffic management with a statistical approach for data mining. Another study of ship’s trajectories is conducted by Hexeberg et al. (2017) for autonomous surface vessels (ASVs). The study aims to predict the next position and time for avoiding collision of the vessels with the use of historical AIS data. The proposed algorithm acquired 30 minutes of vessel trajectory with a good potential. The ship motion prediction with the analysis of sensor data is studied by Li et al. (2016). The authors conducted data cleansing methods with three different methods for the raw sensor data. Moreover, correlation analysis is performed before the training and testing of the neural network. The generated dataset is analysed with the proposed neural network algorithm and an efficient prediction is obtained. Similarly, Wang et al. (2017) focused on the daily ship traffic volume with AIS data. The proposed methods which are Auto-Regressive, Moving Average (ARMA) and Artificial Neural Network (ANN), and hybrid model used for the mining of three different ship types in Shanghai port. The hybrid model provided an efficient traffic volume prediction for cargo, tanker, and passenger vessel types. As the other important topic of maritime transportation, the fuel consumption prediction is studied by different authors. The onboard automation system is used by Coraddu et al. (2017) to investigate the optimisation between fuel consumption and the trim of the vessel. The authors compared three different approaches as white, grey, and black box models. The results showed that the grey box model could be used for optimising the trim in real-time operations. Similarly, Wang et al. (2018) employed a framework for prediction of the fuel consumption with LASSO regression to deal with multiple collinearity problems which could arise when using traditional multiple regression methods. The authors suggested that the main approach would increase the efficient estimation of fuel consumption.

The container terminal management problems are studied with the help of machine learning techniques. Pani et al. (2014) aimed to forecast the late arrivals in a container terminal. The authors discussed an analysis of the data and presented preliminary results with data mining techniques. Furthermore, the approach with a Classification and Regression Trees (CART) model was presented. Likewise, Yu et al. (2018) pointed out that the effective prediction of vessel arrivals would improve
the performance of container terminals. The authors applied CART, RF, and Back-Propagation network (BP) methods to acquire delay or early arrivals of the vessels. The results showed that the Estimated Time of Arrival (ETA) and vessel length were found the most important features in Gangji Container Terminal/China.

**METHODOLOGY**

In this section, the basics of the FCM clustering method and machine learning methods which are used for general model selection and determining clusters in the data set were explained.

FCM clustering algorithm is a well-known and widely used method in clustering studies. The FCM was firstly introduced by Dunn (1973) and improved by Bezdek et al. (1984). The method is based on the minimization objective function:

\[
J_m = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}^m \|x_i - v_j\|^2 ; 1 < m < \infty
\]  

(1)

\(m\): fuzzy partition matrix exponent which is greater than 1

\(u_{ij}\): the degree of membership of \(x_i\) in the cluster \(j\)

\(x_i\): the \(i\)th pattern of \(D\)-dimension data

\(v_j\): \(j\)th cluster centre of the \(D\)-dimension

\(\|x\|\): any norm expressing the similarity between any measured data and the centre

To explain more in detail, the FCM clustering procedure is following as mentioned before (Saxena et al., 2017):

<table>
<thead>
<tr>
<th>Table 1. FCM clustering procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Set up a value of (c) (number of the cluster)</td>
</tr>
<tr>
<td>2. Select initial cluster prototype (V_1, V_2, \ldots, V_c) from (X_i, i=1, 2, \ldots, N)</td>
</tr>
<tr>
<td>3. Compute the distance (|X_i - V_j|) between objects and prototypes</td>
</tr>
<tr>
<td>4. Compute the elements of the fuzzy partition matrix ((i = 1, 2, \ldots, N; j = 1, 2, \ldots, c))</td>
</tr>
<tr>
<td>5. Compute the cluster prototypes ((j = 1, 2, \ldots, c)) (V_j = \sum_{i=1}^{N} u_{ij}^2 x_i \sum_{i=1}^{N} u_{ij}^2)</td>
</tr>
<tr>
<td>6. Stop if the convergence is attained or the number of iterations exceeds a given limit. Otherwise, go to step 3.</td>
</tr>
</tbody>
</table>

Iterative optimization of the objective function is used for fuzzy partitioning, with the update of membership \(u_{ij}\) and the cluster centres \(c_j\):
\[ u_{ij} = \frac{1}{\sum_{k=1}^{c} \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{2(m-1)}} \]  
\[ c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \]  

(2)

The above iteration stops when \( \max_{i,j} \left\{ u_{ij}^{(k+1)} - u_{ij}^{(k)} \right\} \leq \varepsilon \), where \( \varepsilon \) is a termination criterion which is between 0 and 1, whereas \( k \) is the iteration steps. A local minimum or a saddle point of \( J_m \) are converged by this procedure.

The performance of clustering is evaluated with cluster validity functions. In the scope of the proposed methods during the last ten years, FCM could be analysed with two important types of criteria which are based on the fuzzy partition of the sample set and the geometric structure of the sample set. The partition coefficient and the partition entropy are the functions that represent the main idea of validity functions based on fuzzy partition. The studies conducted by Dunn and Bezdek regarding fuzzy partition gives an idea about when the best analysis is achieved. The authors stated that the value of partition coefficient (\( Vpc \)) and the value of partition entropy (\( Vpe \)) may be led to a good interpretation. When the \( Vpc \) gets the maximum value and \( Vpe \) gets the minimum value, the best clustering is achieved (X. Wang et al., 2004).

Briefly, the formula of partition coefficient and partition entropy are the following:

\[ Vpc(U) = \frac{\sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij}^2}{n}, \text{Max}(Vpc) \]  

(3)

\[ Vpe(U) = -\frac{1}{n} \left\{ \sum_{j=1}^{n} \sum_{i=1}^{c} u_{ij} \log u_{ij} \right\}, \text{Min}(Vpe) \]  

(4)

Machine learning is a comprehensive method that computer or machine learns from past experiences and makes predictions. The past experiences are the input data of the model. Generally, machine learning is categorized into three types: (1) supervised, (2) unsupervised, and (3) reinforcement learning. In this study, five machine learning methods are used to make suitable models for operational data analysis.

The basic idea of the MLR is to create a linear equation to model data. To fit data to an equation, the method uses explanatory variables to predict the response variable’s outcome. In other words, MLR is used to calculate the relationship between one dependent variable and two or more independent variables. The form of the multiple regression model is following (James et al., 2013):

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_j X_j + \epsilon \]  

(5)

\( X_j \) represents the \( j \)th predictor, \( \beta_j \) quantifies the association of variable and the response, and \( \epsilon \) is an error.
Ridge regression is a method for analysing multiple regression data that has multicollinearity. The ridge regression aims to reduce standard errors by adding a degree of bias to the model estimates. The least-squares fitting procedures could be formulated as follows with residual sum of squares (RSS):

$$RSS = \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$

(6)

Ridge regression is very similar to the above equation, except that the method penalizes the size of parameter estimates:

$$\sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = RSS + \lambda \sum_{j=1}^{p} \beta_j^2$$

(7)

$$\lambda \geq 0$$ is a tuning parameter and $$\lambda \sum_{j=1}^{p} \beta_j^2$$ is a shrinkage penalty. The tuning parameter server to control the relative impact to generate a different set of coefficient estimates. When the tuning parameter equals zero, the penalty term is neglectable and the ridge regression produces the least squares estimates. On the other hand, when the tuning parameter goes to infinity, the effect of the penalty grows then regression coefficient estimates approaches zero (James et al., 2013).

LASSO regression which is used when the data set has a large number of predictor variables stands for Least Absolute Shrinkage and Selection Operator. Although ridge and LASSO regression are almost similar, the LASSO regression converts the coefficients which cause large variances. Differently, the ridge regression shrinks the coefficients not making them equal to zero. The lasso coefficients, $$\beta_{L}$$ minimize the quantity (James et al., 2013):

$$\sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| = RSS + \lambda \sum_{j=1}^{p} |\beta_j|$$

(8)

As seen in Equations (7) and (8), the ridge and LASSO regression have similar formulations. The term in the LASSO regression $$|\beta_j|$$ is different than the term in the ridge regression $$\beta_j^2$$. In other words, an $$\ell_1$$ in the lasso regression is used instead of $$\ell_2$$ norm. $$\beta$$ as the $$\ell_1$$ norm of a coefficient vector is given by $$\|\beta\|_1 = \sum |\beta_j|$$.

PCR uses the results of Principal Component Analysis (PCA) as new regressors. The output acquired from PCA which is performed on regressors is used as new regressors. PCA has two goals in regression. The first goal is to reduce dimensionality on datasets where the number of predictor variables is too high. The second goal is to deal with collinearities between variables (James et al., 2013).

The PCA is sensitive to data that has not been centred. To cope with this problem, all the variables should be standardised before performing PCA. One of the pre-processing steps of PCA is to normalize
data to get $\mu=0$ and $\sigma=1$. Then, PCA is performed on matrix $X$ through singular value decomposition. So, $X=UDV'$. $D$ which is a diagonal matrix consisting of the number of explanatory variables, $p$.

$$D = \text{diag}[\delta_1, \delta_2, \ldots, \delta_p] = \begin{bmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_p \end{bmatrix}$$ (9)

The $p \times p$ matrix $V$ with eigenvectors as columns is obtained. The matrix $Z = XV$ where each column is one of $p$ principal components is calculated using matrix $V$. By computing in this way, we assure these columns are orthogonal without collinearity. To perform the regression, the $Z$ matrix which contains $r$ or $p$ principal components is used. The coefficients:

$$\beta_Z = (Z'Z)^{-1}Z'y$$ (10)

By using equation (10), we obtain;

$$\hat{y} = Z\beta_Z = (XV)\beta_V = X(V\beta_V) = X\beta_V$$ (11)

$$\beta_X = V\beta_Z$$ (12)

PLS regression performs least squares regression on the predictors which are reduced to a smaller set of uncorrelated components instead of using the original data. The PLS regression gives an advantage in regression studies when the predictors are highly collinear. PLS assumed the predictors different way from multiple regression which takes as fixed. This flexibility makes the PLS more robust to measure uncertainty by calculating the predictors with errors.

The model of PLS regression (Sawatsky et al., 2015):

$$X = TP^T + E$$ (13)

$$Y = UQ^T + F$$ (14)

in which;

The matrices of predictors ($n \times m$ matrix of predictors) and responses ($n \times p$ matrix of response) are $X$ and $Y$. The projections of $X$ and $Y$ are represented as $T$ and $U$ which are $n \times l$ matrices. The orthogonal loading matrices for the projected $X$ and $Y$ scores are represented as $P$ and $Q$. The assumed independent error terms for the predictor and response matrixes are $E$ and $F$.

RF algorithm is an ensemble method that provides some decision trees so that every time random sample of some predictors to split suitable ones from the full set of predictors is selected. The algorithm forces each split to consider only a subset of the predictors. With the help of this process, which is
called decorrelating, the average of final trees is more reliable (James et al., 2013). To evaluate the performance of the machine learning model on a given data set and the distance of each node, the well-known measure is the Mean Squared Error (MSE) which will be small if the predicted responses are close to the true responses. Briefly, the MSE:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left( y_i - \hat{f}(x_i) \right)^2$$

(15)

Moreover, the Mean Absolute Error (MAE) measures the average magnitude of the errors without considering their direction in a set of forecasts. Briefly,

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| = \frac{1}{n} \sum_{i=1}^{n} |e_i|$$

(16)

On the other hand, another evaluation metric is Root Mean Squared Error (RMSE). The RMSE is a scoring rule which is quadratic measures the average magnitude of the error. The differences between prediction and actual observations are calculated to get the square root of the mean of the square of the errors. The general-purpose error metric is;

$$RMSE = \frac{1}{n} \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right|^2$$

(17)

The cross-validation technique is being used for evaluating machine learning models by training several different machine learning models on subsets of the available input data. In k-fold cross-validation, the input data is divided into k subsets of data which is known as folds. The machine learning model is trained all “k-1” of the subsets to evaluate the model on the not used subset for training. The process of training is continued “k” times with a different subset reserved for evaluation each time. This process involves dividing the set of observations into k folds randomly with approximately equal sizes. Then, the mean squared error is computed on the observations in the held-out fold. The k-fold cross-validation:

$$CV(k) = \frac{1}{k} \sum_{i=1}^{k} MSE_i$$

(18)

Support Vector Machine (SVM) is a tool that uses the machine learning theory to achieve maximum predictive accuracy by avoiding overfitting for prediction and classification problems (Jakkula, 2006). SVM aims to separate different classes in the training set to maximise the margin (Cervantes et al., 2020). the kernel method is an alternative use for SVM to enable a researcher to model data with higher dimensions (Huang et al., 2018). The formula of a kernel function:

$$K(x, y) = \langle f(x), f(y) \rangle$$

(19)
MLP can be used for regression and classification problems. The output of the algorithm takes values between 0 and 1. MLP has structures built in the same structure as a nervous system. The main layers of the neurons are (1) input, (2) hidden(s), and (3) output (Agirre-Basurko, 2006). The formula of the network:

\[
Y_k^0 = f_k^0\left(b_k^0 + \sum_{i=1}^{S} w_{ik}^0 y_i^h\right) = f_k^0\left(b_k^0 + \sum_{i=1}^{S} w_{ik}^0 \cdot f_j^h\left(b_j^h + \sum_{j=1}^{N} w_{ji}^h x_j\right)\right), k = 1, \ldots, L
\]  

(20)

**CASE STUDY & DATA OF THE CONTAINER TERMINAL OPERATION**

The container terminal is the gate for the global business environment in a competitive way. The management office of the container terminal has to arrange cargo operations before the vessel’s estimated time of arrival. Due to the high number of containers that are carried by merchant vessels, every operation step has to be optimized to work efficiently. In this manner, the yard and crane automation technologies have become the key elements of the terminals. With the help of integration between software applications and human interfaces, the terminal authority aims to operate the terminal flawlessly. Beginning the berthing of the vessel, the quay cranes, terminal tractors, and port workers are arranged to serve together without any delays. Although pre-planning is done perfectly, the weather conditions would affect the total time of cargo operation. For example, quay cranes are prone to weather conditions such as wind that causes the swing of the cargo and causes a delay in handling time. This weakness takes place when the operation is commenced with mobile handling crane more excessive. To cope with the swing of the cargo/container, the crane operator has to work with less speed.

In the view of weather effects, operational data of container terminal which covers five years is used to explain or predict possible scenarios (windy or calm condition). The dataset consists of vessel name, characteristics of vessel dimensions, container weight and quantities, operation description, the status of the container, air temperature, wind, and humidity. The dataset contains 345,260 rows of the aforementioned information to analyse. The data is acquired directly to the server of the system with the SQL programming language and contains all information without any pre-processing of data. The dataset used in the study is limited to a five-year operation process due to the storage limitation of terminal operating system software. As seen in Figure 1, the numeric data of the study is illustrated with violin plots on a monthly for a five-year period beginning from 2015 to 2019. The categorical data of the study are handling operation description (loading, discharging, and shifting), handling crane names, and container status (loaded or empty). Since the categorical variables affect handling times, all data used for analysing process to acquire significant results. Weather information is gathered from the Turkish State Meteorological Service separately. Furthermore, to cluster the port of calls of vessels, all data was calculated independently to create an order based (752 calls) total variables such as total cargo weight, total cargo move, and total handling time.

In the first step of the study, the dataset of the container port is used for the pre-processing step. The data is analysed to clear any misleading data which may cause the acquisition of wrong model assumptions. To decide which handling time is suitable for data mining, the maritime assumption for mobile gantry crane handling capacity which is 15-25 lifts/h is used (Thoresen, 2014). Following the pre-processing step, the variables for FCM are created as total cargo weight for order, total handling time for order, and total container movement quantity to cluster orders unsupervised approach. Moreover, vessel length and width are used because the dimensions of the vessels could cause differences at cargo handling times. The scope of the clustering is to divide cargo operations in the view of operational handling times which could be affected by various elements. In the third step, the dummy variables and fuzzy cluster membership values are generated as independent variables to analyse the dataset. Then, machine learning methods in Section 3 are used and compared by error
Figure 1 The violin plot of the variables

Figure 2. Flow chart of the study
metrics. At the final step of the study, the weather effects on the container cargo operations are defined as seen in Figure 2.

RESULTS AND DISCUSSION

The study is completed with R studio/R programming language and MS Win 10 Pro environment with Intel quad-core processor and 32 GB memory. The packages of R Studio are used i.e. “ppclust, dplyr, factoextra, cluster, fclust, caret, psych, metrics, e1071, ggplot2, neuralnet, and tidyverse”. All variables are selected differently for the FCM clustering method and machine learning techniques.

Fuzzy C-Means Clustering

In the second phase of the study, order-based data is analysed with the FCM clustering method. In this manner, the possible operation planning solutions and vessel dimensions are defined as the idea of clustering with FCM. The clusters are assigned as two, three, and four for different scenarios. The two-cluster is defined with the idea of the good or bad operation. The third cluster is defined as possible effects on operational planning likewise low, medium, or high performance. The other approach for cluster determination regarding number validation, the container vessels have evolved generations since the mid-1950s. The evolution of container ships could be categorized as early containership, Panamax, Post-Panamax, Neo-Panamax, and very large containership (VLCS). In the dataset of study, there are four generations of container vessels except for the VLCS.

The top and bottom five rows of two-cluster membership degrees matrix (good or bad situation), three-cluster membership degrees matrix (low, medium, or high situation), and four cluster membership degrees matrix (for container vessel evolution generations) are as follows:

Table 2. Membership degree matrix for clusters

<table>
<thead>
<tr>
<th>Order Number</th>
<th>Two-cluster</th>
<th>Three-cluster</th>
<th>Four-cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#1</td>
<td>#2</td>
<td>#1</td>
</tr>
<tr>
<td>1</td>
<td>0.591</td>
<td>0.408</td>
<td>0.109</td>
</tr>
<tr>
<td>2</td>
<td>0.949</td>
<td>0.050</td>
<td>0.030</td>
</tr>
<tr>
<td>3</td>
<td>0.648</td>
<td>0.351</td>
<td>0.167</td>
</tr>
<tr>
<td>4</td>
<td>0.332</td>
<td>0.667</td>
<td>0.051</td>
</tr>
<tr>
<td>5</td>
<td>0.950</td>
<td>0.049</td>
<td>0.022</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>748</td>
<td>0.829</td>
<td>0.170</td>
<td>0.099</td>
</tr>
<tr>
<td>749</td>
<td>0.704</td>
<td>0.295</td>
<td>0.136</td>
</tr>
<tr>
<td>750</td>
<td>0.902</td>
<td>0.097</td>
<td>0.070</td>
</tr>
<tr>
<td>751</td>
<td>0.993</td>
<td>0.006</td>
<td>0.009</td>
</tr>
<tr>
<td>752</td>
<td>0.903</td>
<td>0.096</td>
<td>0.043</td>
</tr>
</tbody>
</table>

The cluster plot is following:

To further evaluate the FCM clustering method on the dataset, the error metrics are the following:

The validation results are used to create a dummy variable for machine learning techniques rather than defining which cluster is fit exactly for the study dataset. As mentioned before, operational differences are defined for the clustering number to analyse possible weather effects on container
handling operations. Fuzzy dummy variables are created by membership degree values as seen in Table 2. On the other hand, one-hot-encoding variables are created with the help of the FCM clustering algorithm which gives exact numbers about cluster assignment as seen in Figure 3.

**The Applications of Machine Learning Techniques**

Machine learning techniques which are MLR, Ridge, LASSO, PCR, PLS, SVR, RF, and MLP regression techniques are applied to the study dataset along with the new variables which are gathered from the FCM clustering algorithm. The different algorithms are selected rather than applying all possible techniques due to the capability of handling problems such as multicollinearity, large variances, bias, and errors. On the other hand, MLR has been selected because of its explanation capability the models are easier than other methods.

The dependent variable which is the handling time of one container and independent variables which are air temperature, rainfall quantity, humidity, wind speed (m/s), operation description, crane
identifications, container weight, and container status (loaded or empty) are divided into two sets as training and test data by 75% and 25%. Furthermore, the correlation test is completed as in Figure 4.

Figure 4. The correlation matrix

At first glance, training data is analysed with MLR without any dummy variables and the RMSE is found as 112.9. Following the application of maritime mobile gantry crane handling capacity requirements on the original data set, the RMSE is found as 26.25 and R-squared is found as 0.005. These metrics are compared with the results of the study to assess the goodness of fit. The different nonlinear approaches for regression are used to acquire a more accurate model on the initial dataset within the scope of RMSE. RF regression has an identical RMSE as 114 despite the increased computational time (three times higher). Similarly, SVM regression generates almost the same RMSE as 118. On the other hand, MLP regression is completed with RMSE as 199.

To validate the results of the machine learning techniques 10-fold Cross-Validation is used, the detailed results are as follows:

To compare the results between MLR and other mentioned machine learning methods, the comparison results of error metrics are following in Table 5. The machine learning methods which are used in this study have the same error metrics for dummy variables. In contrast, although the models which contain one-hot and fuzzy dummy variables have a minor difference for MAE and RMSE, R-Squared (R-Sq) metric increases in every model which contains fuzzy dummy variables. The best model is found as RF regression with two-cluster fuzzy dummy variables.

In summary, the final results show that utilizing FCM clustering methods with machine learning approaches gave a better explanatory solution for container terminal operation analysis. Using fuzzy dummy variables acquired via the FCM clustering technique, the operational data analysis has also proven that wind speed has a negative impact on container handling times. Due to the complicated nature of cargo handling operations, the FCM clustering method is also shown to be useful for analysing container terminal operations for the port of call or order based on an unsupervised approach.

Lastly, since mobile handling cranes have higher carriage capacity than a single fully loaded container, such as 150 metric ton per lift, container weight is found to be unrelated to container handling time. In the sample container terminal dataset, the probable influencing visibility effect of
Table 4. Coefficients of MLR with fuzzy dummy variables as the best model

| Coefficient                                      | Estimate | Std. Error | t value | Pr(>|t|) | Variance inflation factor (VIF) |
|--------------------------------------------------|----------|------------|---------|----------|-------------------------------|
| Intercept                                        | 205.2    | 1.701      | 120.566 | 0.001*** | -                             |
| Humidity                                         | 0.08     | 0.017      | 4.598   | 0.001*** | 1.012                         |
| Wind Speed                                       | 0.76     | 0.193      | 3.924   | 0.001*** | 1.011                         |
| Operation Descr. / Discharge - Shifting on Vessel| 104.6    | 15.028     | 6.958   | 0.001*** | 1.115                         |
| Operation Descr. / Discharge - Shifting to Berth  | 7.88     | 4.012      | 1.964   | 0.049*   |                               |
| Operation Descr. / Loading                       | 20.07    | 0.527      | 38.061  | 0.001*** |                               |
| Crane #2                                         | 7.83     | 0.734      | 10.674  | 0.001*** | 1.013                         |
| Crane #3                                         | 4.81     | 0.682      | 7.044   | 0.001*** |                               |
| Crane #4                                         | 4.15     | 0.668      | 6.213   | 0.001*** |                               |
| Container Status - Loaded                        | -3.83    | 0.714      | -5.371  | 0.001*** | 1.063                         |
| Fuzzy Dummy Cluster Variable #1                  | 8.32     | 1.509      | 5.512   | 0.001*** | 2.513                         |
| Fuzzy Dummy Cluster Variable #2                  | 9.37     | 1.371      | 6.836   | 0.001*** | 2.080                         |
| Fuzzy Dummy Cluster Variable #3                  | 27.57    | 1.219      | 22.612  | 0.001*** | 2.636                         |

Table 5. The error comparison table of machine learning models

<table>
<thead>
<tr>
<th>Number of Cluster</th>
<th>Error Metrics</th>
<th>Type of Variable</th>
<th>MLR</th>
<th>Ridge</th>
<th>LASSO</th>
<th>PCR</th>
<th>PLS</th>
<th>RF</th>
<th>SVM</th>
<th>MLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>RMSE</td>
<td>one-hot</td>
<td>113.8</td>
<td>113.8</td>
<td>113.8</td>
<td>113.8</td>
<td>112.9</td>
<td>117.9</td>
<td>196.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>fuzzy</td>
<td>113.1</td>
<td>113.1</td>
<td>113.1</td>
<td>113.1</td>
<td>112.1</td>
<td>117.9</td>
<td>195.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R-Sq</td>
<td>one-hot</td>
<td>0.009</td>
<td>0.009</td>
<td>0.009</td>
<td>0.008</td>
<td>0.009</td>
<td>0.013</td>
<td>0.011</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fuzzy</td>
<td>0.013</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
<td>0.022</td>
<td>0.011</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>one-hot</td>
<td>83.3</td>
<td>83.3</td>
<td>83.3</td>
<td>83.3</td>
<td>82.8</td>
<td>76.5</td>
<td>117.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>fuzzy</td>
<td>82.9</td>
<td>82.9</td>
<td>82.9</td>
<td>82.9</td>
<td>82.1</td>
<td>75.9</td>
<td>119.5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>RMSE</td>
<td>one-hot</td>
<td>113.1</td>
<td>113.1</td>
<td>113.1</td>
<td>113.2</td>
<td>113.1</td>
<td>112.9</td>
<td>117.5</td>
<td>195.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fuzzy</td>
<td>113.2</td>
<td>113.2</td>
<td>113.2</td>
<td>113.2</td>
<td>112.9</td>
<td>117.9</td>
<td>194.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>R-Sq</td>
<td>one-hot</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.009</td>
<td>0.010</td>
<td>0.011</td>
<td>0.013</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fuzzy</td>
<td>0.015</td>
<td>0.016</td>
<td>0.016</td>
<td>0.015</td>
<td>0.015</td>
<td>0.019</td>
<td>0.021</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>one-hot</td>
<td>82.9</td>
<td>82.9</td>
<td>82.9</td>
<td>82.9</td>
<td>82.8</td>
<td>76.3</td>
<td>120.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>fuzzy</td>
<td>82.9</td>
<td>82.9</td>
<td>82.9</td>
<td>82.9</td>
<td>82.8</td>
<td>76.1</td>
<td>116.3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>RMSE</td>
<td>one-hot</td>
<td>113.5</td>
<td>113.5</td>
<td>113.5</td>
<td>113.5</td>
<td>113.5</td>
<td>112.1</td>
<td>118.4</td>
<td>191.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fuzzy</td>
<td>112.8</td>
<td>112.8</td>
<td>112.8</td>
<td>112.8</td>
<td>112.8</td>
<td>114.2</td>
<td>118.7</td>
<td>190.5</td>
</tr>
<tr>
<td></td>
<td>R-Sq</td>
<td>one-hot</td>
<td>0.009</td>
<td>0.009</td>
<td>0.009</td>
<td>0.008</td>
<td>0.009</td>
<td>0.014</td>
<td>0.012</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fuzzy</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
<td>0.013</td>
<td>0.014</td>
<td>0.019</td>
<td>0.022</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>MAE</td>
<td>one-hot</td>
<td>83.1</td>
<td>83.1</td>
<td>83.1</td>
<td>83.1</td>
<td>83.1</td>
<td>82.2</td>
<td>76.6</td>
<td>119.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fuzzy</td>
<td>82.7</td>
<td>82.7</td>
<td>82.7</td>
<td>82.7</td>
<td>83.1</td>
<td>76.7</td>
<td>117.5</td>
<td></td>
</tr>
</tbody>
</table>
rainfall quantity is seen to be negligible. Future research would aim to enhance models or obtain more meaningful results in the case of rainy operations, which might expand handling times.

CONCLUSION

The study demonstrated the use of FCM clustering and machine learning approaches in container port operations with a historical dataset. The findings of machine learning approaches combined with FCM clustering showed encouraging results in regression tasks, suggesting that they might be utilized to investigate operational time losses further.

This paper identifies proof that the FCM clustering method could be used in maritime port operations with different scopes. For future studies, the R-Squared metric could be increased with different operational data. Moreover, the implementation of the FCM clustering method could be enriched with different unsupervised machine learning clustering methods. Also, enhanced berth planning with estimated port stay time could be developed. Finally, the different prediction and classification algorithms for cargo handling times could be applied to reduce error metrics and increase the explanatory power of the model.

ACKNOWLEDGMENT

The publisher has waived the Open Access Processing fee for this article.
REFERENCES


Üstün Atak, research assistant in maritime transportation engineering, has Oceangoing Watchkeeping Officer competency and Master of Science degree from Maritime Transportation Engineering at ITU (2019). He has experience on different types of vessels such as tanker, container and bulk carriers with multinational seafarers. His research fields are operational analyzes of container terminals, machine learning methods within the scope of maritime transportation and Intelligent Transportation Systems. He currently studies a PhD in the Maritime Transportation Engineering department.

Tolga Kaya is a full-time researcher and lecturer at the department of Management Engineering in Istanbul Technical University. His research areas are consumer modeling, statistical decision making, input-output modeling, multicriteria decision making, machine learning and fuzzy applications in business and management. He has published several papers and presented his research at a number of international conferences in these areas.

Yasin Arslanoğlu works as a professor at the Maritime Faculty of Istanbul Technical University. He is currently responsible for coordination of Maritime Transportation Engineering Graduate Program. He has been teaching environmental courses, maritime chemistry education and machine learning education, etc. He conducted two national projects in researcher positions. He has published 21 papers in reputable international journals, and more than 40 international conference papers in proceedings (h-index 10, more than 325 citations in WoS/WoK). He has been supervising to graduate studies on maritime chemistry, environmental management, machine learning and ship energy efficiency.