*Advanced Logistic Systems – Theory and Practice, 13(1), 27-36, https://doi.org/10.32971/als.2019.009*

# MONITORING OF OPERATIONAL LOGISTIC PROCESSES IN GENERAL CARGO WAREHOUSES USING PREDICTIVE ANALYSIS

ANDREAS NEUBERT[1]

**Abstract:** Due to the different characteristics of the piece goods (e.g. size and weight), they are transported in general cargo warehouses by manually-operated industrial trucks such as forklifts and pallet trucks. Since manual activities are susceptible to possible human error, errors occur in logistical processes in general cargo warehouses. This leads to incorrect loading, stacking and damage to storage equipment and general cargo. It would be possible to reduce costs arising from errors in logistical processes if these errors could be remedied in advance. This paper presents a monitoring procedure for logistical processes in manually-operated general cargo warehouses. This is where predictive analysis is applied. Seven steps are introduced with a view to integrating predictive analysis into the IT infrastructure of general cargo warehouses. These steps are described in detail. The CRISP4BigData model, the SVM data mining algorithm, the data mining tool R, the programming language C++ for the scoring in general cargo warehouses represent the results of this paper. After having created the system and installed it in general cargo warehouses, initial results obtained with this method over a certain time span will be compared with results obtained without this method through manual recording over the same period.
**Keywords:** predictive analysis, data mining, monitoring, general cargo warehouse

## 1. INTRODUCTION

The handling of general cargo in logistics nodes such as general cargo warehouses plays an important role in logistics. General cargo is transported in the warehouses mainly by manually- operated industrial trucks such as forklifts, powered pallet trucks and hand pallet trucks since general cargo has different dimensions, shapes, weights and materials [9] [13]. Errors occur during logistical processes in general cargo warehouses because manual activities are susceptible to possible human error.

According to Lehmann [11], logistical errors are as follows:

- Shortfalls in delivery,
- Incorrect packaging,
- Surplus deliveries,
- Goods failed to arrive, e.g. loss of shipment,
- Wrong goods,
- Goods arrived too late,
- Goods are damaged,
- Shipment to the wrong address and
- Packaging is damaged,
- False/incomplete delivery documents.

The delay in delivery can be due to the fact that the goods or the package had been put down in a storage location that had not been planned beforehand, resulting in long search

[1] PKE Deutschland GmbH
a.neubert@pke-de.com
Germany

times for the package in the goods issue area. The reason for the loss of consignment or the shipment to an incorrect address may be the loading of the package onto the wrong truck.

Predictive analysis makes it possible to discover patterns. If patterns of errors are detected and reported at an early stage, these errors can still be remedied in the general cargo warehouse, thereby saving costs. This paper presents a monitoring procedure for logistical processes in manually-operated general cargo warehouses. Predictive analysis, which also uses data mining inter alia, is employed here.

In the field of data mining, the term score or scoring is used for various contexts. When creating the model, a so-called score function is used to evaluate how well the currently set parameters produce the training data. The process is called scoring in scientific literature [20]. The application of new, unknown data to the model generated by data mining is also called scoring [14]. In this study the term scoring is used in its second sense.

## 2. METHODOLOGY

The methodology describes, in the form of individual steps, what kind of predictive analysis is to be carried out specifically in order to integrate predictive analysis into general cargo warehouses. It also describes how the necessary data for predictive analysis can be obtained.

### 2.1. Proposed methodology

For a predictive analysis, you must first select an approach to realizing predictive analysis in a concept for fast processing of large volumes of unstructured data (step 1). One part of predictive analysis is data mining where a model is built from historical data. A suitable algorithm must be selected for data mining (step 2). In step 3, training and test data are generated from historical data to create a model. To support the creation of the data mining model, a suitable data mining tool must be selected (step 4). Once the model has been created with the data mining tool, it is necessary to consider how the model can be applied to new, unknown data. Since the data mining tool cannot be integrated as a whole into the IT process of general cargo warehouses for runtime and storage reasons, thought must be given to which programming language can be used to read the model and apply the current data to it (scoring) (step 5). The selected programming language is used to create a program that first reads the model and then applies the current process data to the model (step 6). The program is then to be integrated into the IT infrastructure of the general cargo warehouse (step 7).

### 2.2. Definition of the process model

#### *Step 1. Selection of a Data Mining Process Model*

A process model supports the planned flow of a process in phases. Methods are then assigned to these phases. Data mining is used in predictive analysis to build a model from historical data to which new data is then applied.

First, the requirements for a selection of a data mining process model are considered. A planning and management tool must be selected for data mining. The choice of a standard, if possible, is an accepted procedure here. The selected process model must be suitable for the rapid processing of large quantities of unstructured data (big data), as digitization is also making great strides in general cargo warehouses.

What all process models have in common is that a model is first created on the basis of training data. The quality of the model is then checked using test data. The new, unknown data are then applied to the generated model during scoring in order to assess the current situation.

### 2.3. Model creation

#### *Step 2. Definition of the data mining algorithm*

There are different application classes (task areas) for predictive analysis and thus also for data mining, this includes cluster analysis, classification, regression, association analysis, summarization, dependency modeling, change and deviation detection as well as text mining and web mining as specializations [6] [4].

The error situations in the logistical process are to be detected with the help of predictive analysis. For this purpose, it makes sense to classify the current statuses in the general cargo warehouse. Classification classes here include the correct logistical flow and an error situation, whereby this can be further subdivided according to the type of error. The requirement for a data mining algorithm to be selected is therefore the classification of error situations.

#### *Step 3. Generation of training and test data*

In order for a model to be created, training data must first be provided. Like the test data, this data is obtained from the historical data of the process. Each data set is labelled according to its target variables. Here, for example, it is determined whether an error situation or a correct logistical situation exists.

#### *Step 4. Selection of a data mining tool*

Statistical methods and methods of machine learning are used for data mining. Executing these methods manually would be a laborious task.

In order to support the work on data mining, so-called data mining tools are offered. These tools play a supporting role through the automation of processes. The support ranges from data preparation to modelling and validation. The tools are operated in various ways (command line, GUI, icon-based). A data mining tool has to be selected that supports the data mining algorithm obtained in step 2.

#### *Step 5. Creation of the data mining model*

The data mining tool selected in step 4 is used to create the model from the training data. The recognition quality of the model is determined with the test data. After having created the model, the model has to be saved. Here a standard format should be chosen, on the one hand, and a format that is also provided by the data mining tool, on the other hand.

### 2.4. Scoring in the general cargo warehouse

#### *Step 6. Defining a programming language for scoring*

With regard to the application to processes in the general cargo warehouse, the logistical process data must be applied to the model in real time during scoring.

The requirements for a programming language for scoring are a short runtime and its own memory management. The programming language's own memory management reduces the possibility of errors occurring during programming.

***Step 7. Program creation***

Using the defined programming language, a program is created that reads the generated model and then applies the new, unknown data from the process in the general cargo warehouse to the model.

***Step 8. Integration into the IT infrastructure of the general cargo warehouse***

The program created in step 7 is to be integrated into the IT structure of the general cargo warehouse.

## 3. RESULTS AND DISCUSSION

### 3.1. Definition of the process model - Selecting a Data Mining Process Model

Mariscal, Marbán and Fernández [12] examined the data mining process models and methodologies until 2010. This study follows up on their work. Figure 1 shows the development of data mining process models and methodologies up to 2018.

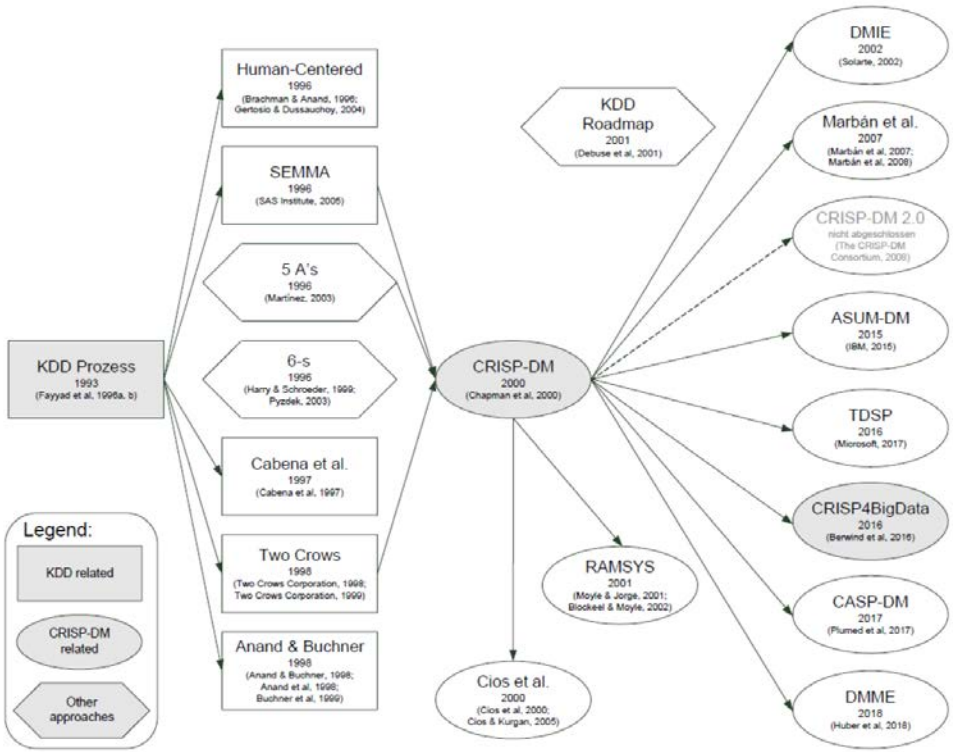


*Figure 1. Development of Data Mining Process Models and Methodologies until 2018 (based on [12])*

Fayyad, Piatetsky-Shapiro and Smyth [7] presented the process of creating patterns in data sets as Knowledge Discovery in Databases (KDD). According to Fayyad et al. [7], KDD is to be understood as "... the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data". After KDD, the Cross-Industry Standard Process for Data Mining (CRISP-DM) was subsequently introduced [3]. CRISP-DM has become a de facto standard [15]. Berwind, Bornschlegl, Kaufmann and Hemmje [1] presented the CRISP4BigData method for Big Data and CRISP-DM (see Figure 2). This method is intended for scientific networks. However, since the method is also suitable for processes in general cargo warehouses, it is proposed for use in this paper.

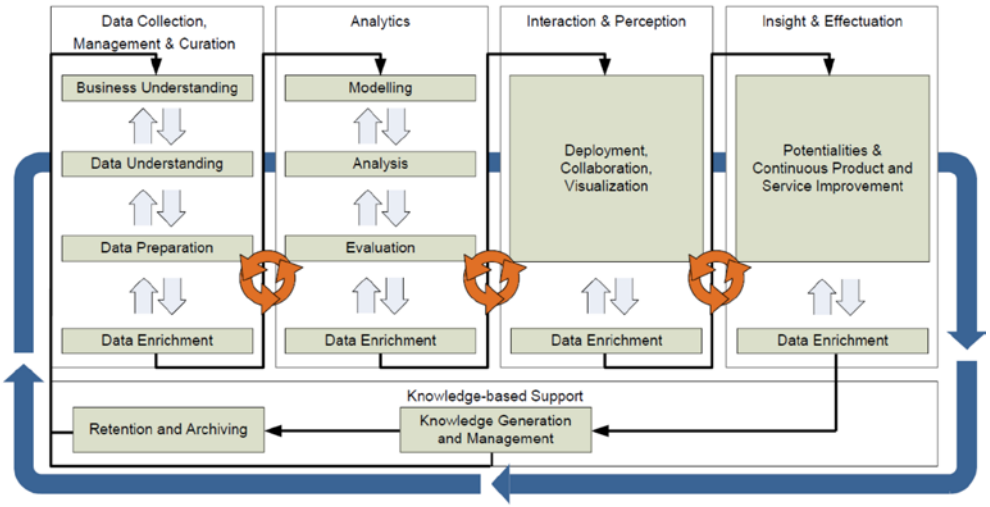


*Figure 2. CRISP4BigData Reference Model Version 1.0 [1]*

### 3.2. Model creation

#### 3.2.1. Definition of the data mining algorithm

Different algorithms are available for classification (see Table I). The Support Vector Machine (SVM) is chosen as the data mining algorithm because it is robust against overfitting. The long runtime during training is negligible, because this time takes place before the actual real-time processing. It is more important that the runtime during scoring is short in order to quickly evaluate the current situation in the general cargo warehouse. SVM is a method based on machine learning. SVM can be used as a classifier and also as a regressor. So-called support vectors are calculated from the input data. The vectors in the data space are separated by a hyperplane.

*Table I.*

*Data Mining Algorithms (based on [10]*

| **Algorithm** | **Description** | **Advantages** | **Disadvantages** |
|---|---|---|---|
| Decision tree | Partitions the data into smaller subsets. | A normalization of the predictors is not necessary. | Tends to overfit. Choosing the right parameters can be difficult. |
| Rule induction | Models the relationship between input and output using IF/THEN rules. | Easy to deploy in almost any tool and applications. | Splits the data set in a straight line. |
| k-Nearest neighbour | Each new unknown data point is compared with a similar known data point in the training set. | Can handle missing attributes in an unknown data set. Works with nonlinear relationships. | The runtime and storage requirements for provision are complex. |
| Naive Bayes | Prediction of the output class according to Bayes' theorem (probability). | The time required for modeling and deployment is minimal. | The training set must be a representative sample of the population. |
| Neural Networks | A model inspired by the biological nervous system. | Good at modeling nonlinear relationships. Fast response time in deployment. | Requires preprocessing of data. Missing attributes cannot be handled. |
| SVM | Essentially an algorithm to detect boundaries (hyperplanes) | Very robust against overfitting. Good at handling nonlinear relationships. | Computational performance during training phase can be slow. |

### 3.2.2. Generation of training and test data

For the generation of training and test data, the data of a shipment tracking system within the general cargo warehouse is suitable, which is used to document transfers of liability, damages and security in general cargo warehouses. In a tracking and tracing system, the data of the logistical process have to be saved for later research. This archived data can be used to generate training and test data. The installation of a tracking system in a general cargo warehouse is the state-of- the-art today.

For labeling, each data record must be classified to determine whether it is an error situation or a correct situation. The data of a consignment tracking system can also be used for this purpose. The documentation of searches carried out is a demand made by the works council of the operators of general cargo warehouses. Whenever a search is carried out, a logistical error has occurred in the past. Either a package is not in the planned storage location or a damage inquiry must be carried out. The datasets that are searched for can then be marked as error situations.

### 3.2.3. Selecting a data mining tool

A market study was carried out to select a suitable data mining tool.

Requirements for a suitable data mining tool are a large number of functions it can provide as well as good options of distributing the results (model). The software environment for statistical calculations and graphics R was selected as the data mining tool [19]. This tool is operated via the command line. The creation of R code is supported by the integrated development environment (IDE) RStudio [18]. R provides a large number of functions with 13,896 packages [5]. There are also many distribution/export options. Thus, source code, a bytecode compiled program, RDA, PMML, POJO, PFA, ONNX and MLeap are provided.

### 3.2.4. Creating the data mining model

With the data mining tool R the model can be created from the training data. The recognition quality of the model is determined with the test data. After having created the model, it has to be saved. The Predictive Model Markup Language (PMML) format [8] is suitable for storage here, which is a standard file format for saving models. The data mining tool R offers an export option to PMML.

## 3.3. Scoring in the general cargo warehouse

### 3.3.1. Definition of a programming language for scoring

The following benchmarks for data mining and other scientific calculations are worth mentioning:

- SciMark 2.0, Java benchmark for scientific and numerical calculations [16] and
- NU-MineBench, Data Mining Benchmark Suite [17].

In SciMark 2.0 the algorithms were implemented in the Java and C programming languages. In the benchmark NU-MineBench, the algorithms were implemented in C and C++. Table II compares the properties of the programming languages.

*Table II.*

*Properties of Programming Languages*

| Programming language | Paradigm | Code | Memory management |
|---|---|---|---|
| C | imperative | Machine language | No |
| C++ | OO | Machine language | Yes |
| Java | OO | Byte code | Yes |

Programming language C is not suitable for scoring since programming language C does not have its own memory management. This leads to a higher susceptibility to errors. The programming languages C++ and Java, on the other hand, have their

own memory management. In the Java programming language, the source code is translated into a so-called byte code, which is converted by an interpreter program into machine language, which is then executed. The platform independence of Java is thus realized. The intermediate step via the interpreter program extends the runtime of the program in comparison to the direct generation of machine language as in the case of C++. For performance reasons, the programming language C++ is selected for scoring.

C++ is therefore chosen as the programming language. Chang and Lin [2] have provided a library LIBSVM for SVMs. This library LIBSVM is available for use with the data mining tool R in package e1071.

#### 3.3.2. Program creation

With the programming language C++ a program is to be generated, which reads the created model and applies the current data from the process by means of the library LIBSVM to the model. If an error situation is detected in the logistical process, it must be reported to the warehouse clerk.

#### 3.3.3. Integration into the IT infrastructure of the general cargo hall

The program created must be integrated into the IT infrastructure of the general cargo warehouse. Data interfaces from various systems, such as the forklift guidance system or the shipment tracking system, must be connected here.

## 4. CONCLUSIONS

The study described necessary steps for the monitoring logistical processes in a general cargo warehouse using predictive analysis.

The CRISP4BigData model was chosen as the process model suitable for big data processing.

The SVM data mining algorithm proved to be best for model creation. A method based on a shipment tracking system installed in the general cargo warehouse was proposed for the necessary generation of training and test data. The data mining tool R was selected to support model creation.

For the scoring in the general cargo warehouse, the programming language C++ was defined, which is then used to create a program that reads the model and applies the current data to the model. Interfaces for the program to various IT systems in the general cargo warehouse were discussed.

The definition of suitable data interfaces to the necessary IT systems in the general cargo warehouse is the subject of further research. The European General Data Protection Regulation (GDPR), which will come into force in May 2018, sets specific requirements for the processing of personal data. Since there are people working in general cargo warehouses, appropriate measures must be taken to protect personal data. There is a need for research in this area in order to develop suitable protective measures such as the anonymisation of personal data.

**References**


[1] Berwind, K., Bornschlegl, M., Kaufmann, M., & Hemmje, M. (2016). Towards a Cross Industry Standard Process to support Big Data Applications in Virtual Research Environments (Long Paper). In Conference: *Collaborative European Research Conference* (CERC) (pp. 82–91). Cork, Ireland: Cork Institute of Technology.

[2] Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1-27. https://doi.org/10.1145/1961189.1961199

[3] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 - Step-by-step data mining guide*. Retrieved from https://www.the-modeling-agency.com/crisp-dm.pdf

[4] Cleve, J., & Lämmel, U. (2016). *Data Mining* (2nd ed.). Berlin; Boston: De Gruyter Oldenbourg. https://doi.org/10.1515/9783110456776

[5] Comprehensive R Archive Network (CRAN). (1997). *CRAN - Contributed Packages*. Retrieved March 15, 2019, from https://ftp.fau.de/cran/web/packages/index.html

[6] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, *17*(3), 37–54.

[7] Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. (1996). From data mining to knowledge discovery: an overview. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth & R. Uthurusamy (ed.), *Advances in knowledge discovery and data mining* (pp. 1-34). American Association for Artificial Intelligence.

[8] Guazzelli, A., Lin, W. C., & Jena, T. (2012). PM*ML in Action: Unleashing the Power of Open Standards for Data Mining and Predictive Analytics* (2nd ed.). CA: CreateSpace Independent Publishing Platform.

[9] Klaus, P., Krieger, W., & Krupp, M. (2012). *Gabler Lexikon Logistik: Management logistischer Netzwerke und Flüsse*. Wiesbaden: Gabler Verlag. https://doi.org/10.1007/978-3-8349-7172-2

[10] Kotu, V., & Deshpande, B. (2014). *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*. E-Book. Waltham: Elsevier. https://doi.org/10.1016/B978-0-12-801460-8.00013-6

[11] Lehmann, U. (2007). *Vortrag „Logistische Fehler“*. Beitrag präsentiert auf dem/der DGQ Regionalkreis Berlin, Deutsche Gesellschaft für Qualität (DGQ), Berlin, Deutschland. Retrieved March 16, 2019, from https://www.dgq.de/dateien/Vortrag_Lehmann_DGQ_Berlin_16_05_2007_Logistische_Fehler.pdf

[12] Mariscal, G., Marbán, O., & Fernández, C. (2010). *A survey of data mining and knowledge discovery process models and methodologies.* The Knowledge Engineering Review, Cambridge University Press. https://doi.org/10.1017/S0269888910000032

[13] Martin, H. (2016). *Transport- und Lagerlogistik: Systematik, Planung, Einsatz und Wirtschaftlichkeit* (10th ed.). E-Book. Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-14552-1

[14] Obenshain, M. K. (2004). Application of Data Mining Techniques to Healthcare Data. *Infection Control & Hospital Epidemiology*, *25*(08), 690–695. https://doi.org/10.1086/502460

[15] Ponce, J., & Karahoca, A. (2009). *Data mining and knowledge discovery in real life applications*. Croatia: In-Teh. https://doi.org/10.5772/97

[16] Pozo, R., & Miller, B. (2004, March 31). *Java SciMark 2.0*. Retrieved March 9, 2019, from https://math.nist.gov/scimark2/

[17] Robert R. McCormick School of Engineering and Applied Science, Northwestern University. (2015, February 22). *NU-MineBench - CUCIS.* Retrieved March 9, 2019, from http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html

[18] RStudio, Inc. (2018, October 12). *RStudio - RStudio.* Retrieved March 15, 2019, from https://www.rstudio.com/products/rstudio/

[19] R Core Team (2018). R: *A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Retrieved March 16, 2019, from https://www.R-project.org/.

[20] Wang, S., Chaovalitwongse, W. A., & Seref, O. (2011). Operations Research in Data Mining. *Wiley Encyclopedia of Operations Research and Management Science*, p 9. https://doi.org/10.1002/9780470400531.eorms0596