

Improve the Orders Picking in e-Commerce by Using WMS Data and BigData Analysis

Augustyn Lorenc

Assistant Professor
Cracow University of Technology
Poland

Aurelija Burinskiene

Associated Professor and senior researcher
Vilnius Gediminas Technical University
Lithuania

The primary purpose of the research is the improvement of the orders picking process without additional investments for the software, employees, tool and inventories. For problem-solving, the data about picking is exported and preprocessed from WMS. The BigData analysis and product clustering in Tableau software is delivered using the data, where the Product Allocation Problem (PAP) is solved. Picking time for reference scenario and new analysed one is calculated and compared.

The presented research proves that standard data collected by WMS could be used for solving PAP for the reduction of total picking time. The method delivered by authors could be in a typical warehouse, where forklifts and employees do the order picking process. The plan after an upgrade could be used for automatic picking, and implemented WMS. For BigData analysis, Tableau is connected to WMS database. Such solution could be used for everyday analysis and planning the allocation of products. The presented method is easy to use; there is no need to invest in expensive software and automation of the picking process to achieve the high performance of the orders picking process. However, its application allows the increase of efficiency rates. Storekeepers can select more products at the same time. The presented research is original because of using simple methods and analysis of specific data, which until now are only used to calculate employee performance indicators.

Keywords: Product Allocation Problem (PAP), BigData analysis, Tableau analysis, clustering, the effectiveness of orders picking, e-commerce, warehouse logistics.

1. INTRODUCTION

Nowadays, many support systems are used in distribution centres (DC) and warehouses. It is impossible to stock products and picks orders with high effectiveness without Warehouse Management System (WMS). It is also impossible to manage deliveries without Yard Management System (YMS). Distribution centres have systems supporting almost every internal process. One of the main problems in the XXI century is data overgrowth. Managers have many reports and statistics, but very often, this data is not useful as they need to spend a lot of time for analysing the data, and the results of this analysis are out-of-date. Effectiveness of the picking process also depends on the warehouse size. This criterion determines the product storage policy (Lorenc and Lerher, 2019). Managers expect arrangement, easy to understand results in real-time. Thanks to the development of IT systems, it is possible to connect to the different data sources and combine data for review and deliver BigData analysis in Tableau, Power BI, ClickView. Analysing of the BigData is the response for the Warehouse 4.0 ideology. Modern development of the warehousing

process, known as Warehousing 4.0 requires the simultaneous treatment of production and warehouse infrastructure, transport-warehouse technology and warehouse management systems (Lerher, 2018).

In the presented paper, the available data from WMS is combined with statistics about products and the warehouse layout to improve the efficiency of the picking process. Such revision of data allows noticing the products that are not correctly placed in the warehouse. So, the first possible way to improve the effectiveness of the warehouse is to solve Product Allocation Problem (PAP).

The data from the WMS system is cleaned from useless information. The client's orders are formed according to PAP logic. The product classification model is built using Tableau software. For scenario evaluation Reference Variant (current scenario) with Analysed Variant (proposed scenario) is compared.

2. LITERATURE REVIEW

The term "BigData" is treated as a large data set that can be revised in a computational way to identify patterns, dynamics. This kind of data is hard to analyse and revise without analytical tools. Recently, a large amount of data, which is not always unstructured, is accumulated (Papadopoulos *et al.*, 2017).

But BigData is not just about the amount of data. The author argues that not only volume (a large number of data lines) but also the speed of data generation and

Received: August 2020, Accepted: November 2020

Correspondence to: PhD Augustyn Lorenc

Cracow University of Technology

al. Jana Pawla II 37, 31-864 Cracow, Poland

E-mail: alorenc@pk.edu.pl

doi: 10.5937/fme2101233L

© Faculty of Mechanical Engineering, Belgrade. All rights reserved

FME Transactions (2021) 49, 233-243 233

diversity of data collected from various sources and data formats are of great importance. Even a small dataset created by multiple sources and formats and frequently updated is also considered large due to its diversity and speed. The revision and analysis of such a large amount of data, which is rich in content, can change the processes of decision making. The authors Chongwatpol and Chan (2015) describe a case study for which a large dynamic dataset is used to refine operational decision making performance and efficiency improvement.

The research of Chongwatpol and Chan (2015) is not staying on the term of BigData but shows that the analytics of data brings alternative ways to report business issues. Analytics improve outcomes by increasing the visibility of operations and help identify opportunities for the change of order picking process (Prasad *et al.*, 2018) and also leads for the transformation of other processes of the company (Papadopoulos *et al.*, 2017). One example, of such operations, is demand management, which synchronises supply and demand, increase flexibility, and reduce variability in e-commerce business (Croxtton *et al.*, 2002). Authors gave an example, which can increase incomes in the company by 9.7% with provided BigData-driven forecasting. The researchers point out that demand management connects well BigData and electronic commerce, and they both can deliver significant benefits to the e-commerce company.

Although BigData is beneficial, there are not many research papers for warehouse management.

Even though empirical BigData research is not yet widespread in the operations area, analytical techniques required for BigData organisation and the value this brings to the business is well investigated (Matthias *et al.*, 2017). The critical aspect of value creation is dependant on the company's ability to take advantage of advanced analytical techniques (Arunachalam *et al.*, 2018). In the last decade, there was a dramatic increase of the implementation of information technologies for supply chain operations (in particular, systems for enterprise resource planning, their records, the status of customers orders). The enlarged volume of data results in BigData analysis.

Over a half-century, the BigData analysis evolution starting from advanced statistical and data mining techniques. There are some examples of the application of advanced statistical methods, such as probability analysis, multiple correlations, Markov chains, martingales, decision support techniques, asymptotics and measurements. Most

popular early works focused on productivity measures, one such researcher is Egner (1973).

Data mining techniques highlighted in 1980 by promoting clustering, modelling, outlier detection, prediction and decision tree setup. Herein, data mining is searching for patterns and relations between the variables of the data and providing logical and theoretical descriptions of constructed link, and models (Waller and Fawcett, 2013). In early phase modelling technique is pre-dominant, where are papers published on warehouse management topic by Khan (1984) and Malmberg *et al.* (1986).

A group of separate tools became known as business intelligence. First-generation business intelligence (BI) was focusing on business understanding and fact connected decisions instead of intuition. The first-generation BI was evident in other disciplines in 1990, but in warehouse management research, this system was highlighted one decade later. Second-generation BI was connecting internal and external data of the company, and for that, more powerful tools were required (Davenport, 2013).

BigData analytical tools were pointed out when it became evident that BI systems are not sufficient to compound information delivered from different data points (Arunachalam *et al.*, 2018).

Warehouse network study requires big data set of randomly selected demand of clients, operations of the warehouse, and freight transportation. There is mixed-integer nonlinear programming (MINLP) model intended for warehouse place selection, which is utilising such data. Besides, to this, capacity forecasting model is delivered to support warehouse managers, which is investigating demand data (Hazen *et al.*, 2018). In the recent decade, the number of papers on BigData analysis in the field of warehouse management is continually growing. Most researchers are investigating operational performance, and some of them deliver studies for the functional area of warehouse management. BigData analytics is applicable for the revision of strategical, tactical and operational layers in warehouse management. Techniques for strategical analytics are used for designing warehouse network, methods for other two types of analytics – for scheduling and resource planning (Addo-Tenkorang and Helo, 2016). In the table below, the authors highlight the publications corresponding to this research field (see table 1).

Table 1. Literature review on the evolution of BigData analysis in warehouse management research field

Aspects	1970	1980	1990	2000	2010
Advanced statistical techniques	Crisp (1968); Egner (1973)	Malmberg <i>et al.</i> (1986)	Mineo and Plaia (1998)	Fang and Lin (2006)	Salah <i>et al.</i> (2011)
Data mining techniques		Khan (1984); Malmberg <i>et al.</i> (1986)	Hendrickson (1999)	Han and Mei (2004); Jiang and Wang (2005)	Zezzatti and Ochoa (2012)
First-generation BI				Lawrence <i>et al.</i> (2005)	
Second-generation BI					Weberschlager (2013); Singh (2016)
BigData analytics BDA					Bodden-Streubuhr (2016); Baker <i>et al.</i> (2017)

Table 2. The statistical analysis of papers (2010-2020)

Number of publications	In title: Operation Management		
	63		
	BigData & supply chain management	BigData & Logistics management	BigData & warehouse management
	271	15	20
%	100%	5,5%	7%

Analytics has been used for many years in operations research. There are three types of analytics: prescriptive, predictive and descriptive.

Prescriptive analytics uses optimisation and decision support techniques. It is rather complicated, but large enterprises are using it for schedule optimisation, inventory decisions and performance simulation (Tiwari et al., 2018).

Large scale optimisation is used for analysing multi-objects, revising alternatives and visualising. These provide accuracy in planning as gathering compressive sampling (CS) in signal processing and sensor systems. For finding out trends and patterns when the complexity of data, which is in high dimensional, is simplified by involving principal component analysis (PCA), in finding clusters for more than one dimension by taking subspace clustering technique.

Predictive analytics is used to forecast or predict the future. It well identifies trends or explore patterns and discovers reasons by taking statistical data analysis, events simulation or programming algorithms. Logistics predictive analytics is using both qualitative and quantitative methods to predict the future behaviour based on historical data of storage of products, the volume of flow, the costs of operations and the fulfilment of clients' orders (Waller and Fawcett, 2013).

The role of an analytical tool is especially important as it must balance each element in the framework. For the development of the analytical tool, companies select one product-lines or product-families for analysis (Pettit et al., 2013). After all, this is delivered, the post-diffusion process of predictive analysis involves acceptance, routinisation and assimilation. The acceptance of predictive analysis concerns how well a company perceives it. Routinisation shows the adjustment of company governance to adjust to predictive analytics. And finally, assimilation measures the diffusion of analytics into business processes (Gunasekaran et al., 2017).

Descriptive analytics is mainly used to illustrate, provide reports and insights by using historical or real-time data.

BigData in warehouse management employs similar statistical methods to get sophisticated data analysis. These methods are specific ones, such as the Wilcoxon-Mann-Whitney test, linear least squares method (OLS), the conjugate gradient on the standard equations (CGNR). They are used for data comparison and modelling. One type of techniques is a simulation, which is an important part of prescriptive and predictive analytics. Simulation helps business to predict warehouse and resources capacity needs, improve utilisation, and service.

BigData analytics is a critical factor for all types of operations. Authors also mention various topic connected studies (Choi et al., 2018). To be more precise, the

authors select three subtopics. Authors use Publish and Perish tool version 6 to revise Google Scholars papers matching the critical word "big data" and the title words "operation management", "supply chain management", "logistics management", and "warehouse management". The summary of this revision is provided in table 2.

The papers dedicated for operations management are published from the year 2010; supply chain management – from the year 2012, for logistics management – from 2014, and for warehouse management – from 2016. The trend shows that late publications are growing faster under the last category.

Papers falling under logistics management category are 5.5%, and most of them cover topics of city logistics, transport capacity and 3PL services.

The analysis shows that 7% of papers are dedicated to the application of BigData for warehouse management. Authors analyse Warehouse 4.0 operations, WMS adoption and the implementation of the Internet of things (IoT) as ones that generates real-time data. Still, these analyses do not cover orders picking from an operations perspective neither from e-commerce case application. In the papers placed under the last section of Table 2, WMS is treated as a system generating BigData. Of course, warehouse employees a lot of resources and is the main centre for business operations. That is why the application of BDA in warehouse management is crucial.

BigData tools are enforcing value generation and bringing competitive advantage (Wamba *et al.*, 2018). Insights present from the analysis of vast amounts of data helps to make an evidence-linked decision in most warehouse functions, and actions initiated after these decisions improve the overall e-commerce system performance. Also, warehouse managers have to develop capabilities connected to analytics and technical part as key success factors and build a data-driven organisation. These data analytics services have to be provided for order picking operations in e-commerce warehouse (Dremel *et al.*, 2017).

The massive data generation influences the collection and analysis of this data and results in more complex analysis. The steps are taken for end-result start from (1) BigData generation, (2) BigData acquisition, (3) BigData storage, and (4) BigData analysis. (1) Data could be generated by the WMS system, video cameras, tags, smart devices, RFID data captured by sensors, etc.. BigData processing could be initiated in batches, real-time or by interactive actions when order pickers enter the data. (2) Data acquisition is the extraction of insights from data. Herein, multivariate statistical analysis or Bayesian statistics are applied as powerful techniques. Finally, the BigData framework could help to determine the warehouse events that not efficient. (3) The pre-final step is BigData storage,

which incorporates data into a database or platform with multi-data sources. The quality of the data depends on infrastructure (i.e. good architecture of system) that helps warehouse managers to organise the collection of the data, compilation, holding, transferring and distribution leading to the visibility of the output as quickly as possible (Wamba *et al.*, 2018). (4) The final step is BigData analysis. It could require data analysis, exploration and visual representation (i.e. good delivery of BigData analytics) (Pettit *et al.*, 2013). The most useful are several types of BigData analysis: clustering, segmentation, data development, and visualisation. For further study, multiple methodological approaches could be used (Choi *et al.*, 2018).

When BigData is considered, the clustering algorithm is used the data analysis. The most used clustering method is K-means clustering, which is a centroid approach to distribute various inputs among k clusters in a balanced way. Other ways of grouping are like balance groups based on different parameters, such as distance or density (Chan *et al.*, 2016). The focus of such an analysis is a significant reduction in operational costs. The study below will focus on advanced descriptive (e.g. clustering) and prescriptive analytics and decisions which are proven by evidence.

3. WAREHOUSE DESCRIPTION

The analysed case is based on real data from e-commerce company selling the home appliances, electronics, computers, and audio devices. The company is located in Poland. The data comes from the busiest period in the year - from September to January. Most orders are shipped during this period. Often the basket of products is small. The number of orders and the type of products are not easy to predict.

The warehouse is divided in two areas – marked as P for pallet unit stored in pallet rack (1611 locations) and marked as T for small products stored in shelf racks (10742 locations). In the analysed period the 25 191 unique products are picked (picked in area P 12095, selected in area T 16889). The 106651 clients' orders are collected (orders processed in the area P: 54767 and orders selected in area T: 83106). The demand representing the number of orders per day is presented in Figure 1.

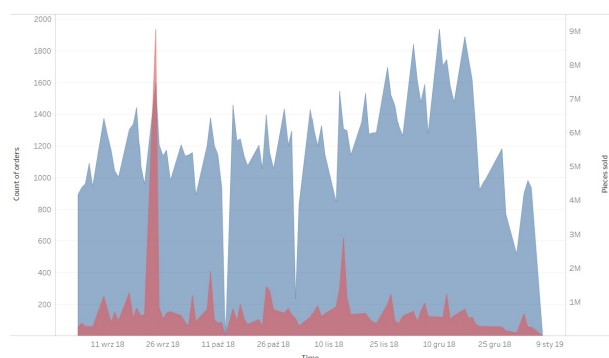


Figure 1. Sales during the period

For picking products, the 70 multi picking trolleys, and 20 standard forklifts are used. There are ten gates allocated for receiving and shipping products. The

warehouse layout consists of pallet racks, shelves, racks and gates, its autoCad version is presented in Figure 2.

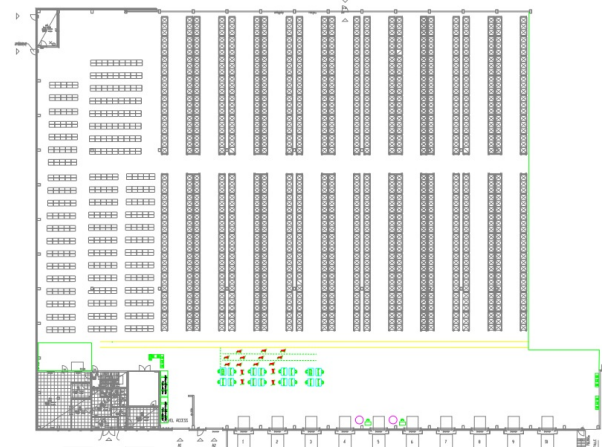


Figure 2. Warehouse layout

The picking process looks like this:

- the product picking is delivered by using for mobile scanning terminals, which suggest the collection path,
- the terminal indicates, among other locations, goods, quantity and requires the confirmation of location barcode placed on the rack and the barcode of a unique product,
- products are placed into the proper level of multi picking trolley, which has Serial Shipping Container Code (SSCC),
- after picking off all products from the list, the storekeeper goes to the shipping area, where he packs the products and print the shipping letter.

The picking process requires optimisation for allocating the optimal number of products. Currently, due to the lack of correct dimensions of goods in the system, the picking tasks exceed the capacity of the transport medium.

The warehouse is working in two shifts. The storekeepers must complete about 80% of the orders before the shipping time (about 4 p.m.), and the second shifts complete orders before 10 p.m. Figure 3, presents the number of picked products and requests during the day.

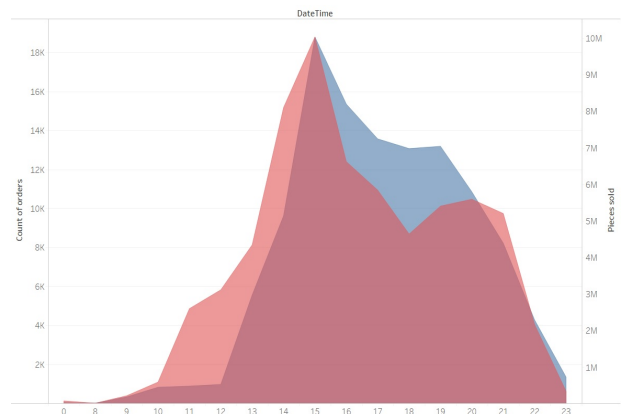


Figure 3. Day activity in the analysed period (blue colour – the number of orders, red ones – the number of products)

The sample of row data is exported from WMS and presented in table 3.

Table 3. Row data exported from the WMS

DateTime	Location	ProductID	Pieces	Pieces Available	SSCC	Order Number	Storekeeper
21.11.2018 22:24	P018-210-10	C555103	-1	0		M3700-1	D.P
21.11.2018 22:24	OUT-WZ-MM	C555103	1	0	93900031797	M3700-1	D.P
21.11.2018 22:23	OUT-WZ-MM	C555103	-28	0	93900031797	M3700-1	D.P
21.11.2018 22:23	OUT-WZ-MM	C555103	49	0	93900031797	M3700-1	D.P
21.11.2018 22:23	P011-360-00	C555103	-21	0		M3700-1	D.P
21.11.2018 22:19	OUT-WZ-MM	C555103	28	0	93900031797	M3700-1	D.P
21.11.2018 22:19	RAMPA7	C555103	-28	0		M3700-1	D.P
21.11.2018 22:17	P003-060-20	C390997	-27	-14		M3700-1	D.P
21.11.2018 22:17	OUT-WZ-MM	C390997	-7	0	93900031797	M3700-1	D.P
21.11.2018 22:17	P003-060-20	C390997	14	14		M3700-1	D.P
21.11.2018 22:17	OUT-WZ-MM	C390997	20	0	93900031797	M3700-1	D.P
21.11.2018 22:17	P003-240-20	C390997	-7	0		M3700-1	D.P
21.11.2018 22:17	OUT-WZ-MM	C390997	7	0	93900031797	M3700-1	D.P

Table 4. Sample of cleared data prepared for further analysis

DateTime	OrderNumber	Location	ProductID	Pieces	Storekeeper
03.09.2018 10:04	10/FM20825/00-1	P006-090-10	C3901323	1	O. A
03.09.2018 10:05	10/FM19575/00-1	P004-390-20	C0055016	3	O. A
03.09.2018 10:06	10/FM19575/00-1	P003-120-20	C2779116	1	O. A
03.09.2018 10:06	10/FM20825/00-1	T015-010-00	C8317800	1	O. A
03.09.2018 13:20	40/FM19428/00-1	P014-090-10	C0409116	1	N. D
03.09.2018 13:21	40/FM20547/00-1	P011-250-00	C0459152	8	N. D
03.09.2018 13:23	40/FM26506/00-1	P005-400-00	C0459138	2	N. D
03.09.2018 13:25	40/FM20415/01-1	P002-070-20	C7722242	15	N. D
03.09.2018 13:27	10/FM17347/01-1	P005-050-10	C0180640	1	O. A
03.09.2018 13:28	40/FM19428/00-1	T020-080-20	C0191998	1	N. D
03.09.2018 13:28	40/FM26362/01-1	P016-020-00	C0409992	1	P. S
03.09.2018 13:29	40/FM18707/00-1	T068-080-10	C4310736	3	N. D
03.09.2018 13:29	40/FM26362/01-1	P016-040-00	C0409648	1	P. S
03.09.2018 13:29	40/FM26362/01-1	P017-170-00	C0459223	2	P. S

Row data exported from the system are hard to understand and hard to analyses. The WMS shows every operation. The in case, storekeeper gets the task to get six pieces of item from the rack, puts places them into containers and leaves four units in the rack.

Because of this, the export data are preprocessed and cleared into more comfortable understandable form (which is presented in table 4).

The data presented in table 4 is used for further analysis.

3.1 Methodology

The research methodology is present in Figure 4. For the research, the two variants are defined. The Reference Variant presents the current state and processes in the warehouse. The Analysed Variant offers the case with relocated products.

3.2 Warehouse layout and topology

For visualisation purposes, the coordinates for every location (i.e. its ID) are identified. The warehouse plan in AutoCad format (Figure 2) is flattened to JPG file. In the warehouse, the location of each rack has XY coordinates, which are entered into the JPG file. The method of calculating these coordinates is presented in Figure 5.

All useless elements are deleted from the plan of the warehouse. Just rack and shelves are presented.

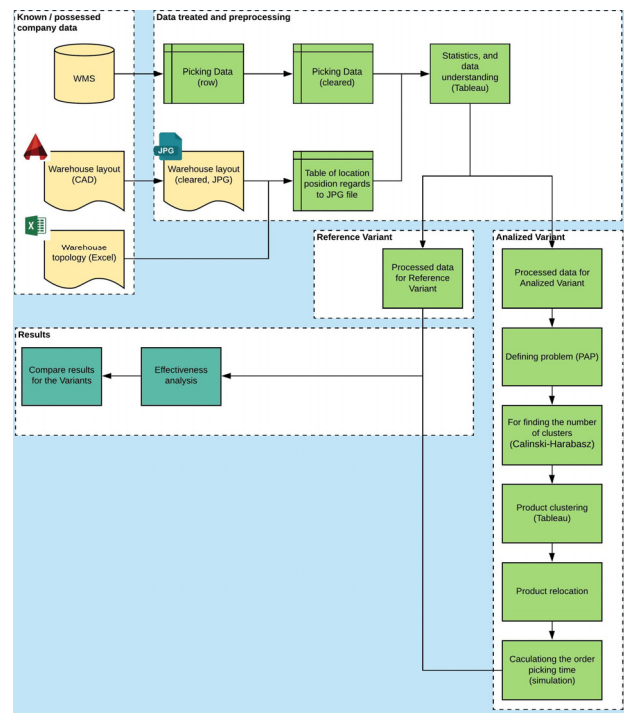


Figure 4. Diagram of the used methodology

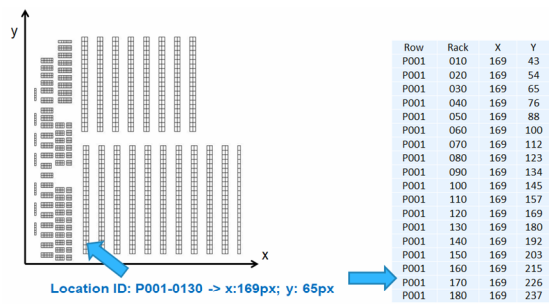


Figure 5. The method of calculating the coordinates for location ID

3.3 Data understanding

The first step of data analysis is oriented to investigate the demand per each product. The number of requests per each product ID is presented in Figure 6.

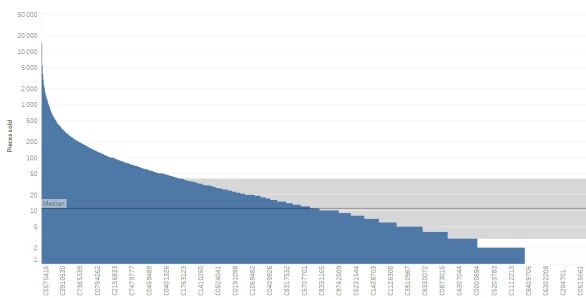


Figure 6. Demand for each product in the analysed period

In the analysed period, many products are demanded once or few times per period. It could mean that:

- for those products, the demand is deficient, and this product should be removed from the offer,
- these products are ordered just for a specific client, and warehouse treat them as cross-docking products, because of that this product should not be analysed and clustered,
- the product ID is changed, but the product is the same, this situation is noticed when the producer changes the package or makes little modification of the product,
- the products with different ID are similar, for example, these products have different colours, but the key parameters and functionalities are the same, so these products should be treated as duplicates and should get one product ID.

3.4 Clustering

For clustering, the 25191 inputs are taken. The criteria for each product are defined accordingly:

- the number of orders in which product occurs,
- the number of items, the difference between actual,
- the number of pieces demanded.

For finding the number of clusters, the Calinski-Harabasz criterion is used (1).

$$CL = \frac{SS_B}{SS_W} \cdot \frac{N-k}{k-1} \quad (1)$$

The data diagnostics for the generated Machine Learning model is presented below:

- between-group Sum of Squares: 48.674,
- within-group Sum of Squares: 5.0912,
- total Sum of Squares: 53.765.

For ML classification model, the algorithm of k-means is used. For a given number of clusters k , the algorithm partitions the data into k clusters. Each cluster has a centre (centroid) that represents the mean data among all points in that cluster. K-means algorithm locates centres through an iterative procedure that minimises distances between each point in the cluster and the defined centre of it.

The centre of a cluster needs to be defined first. In the beginning, the method selects a variable whose mean value is used as a threshold by dividing the data into two parts. The defined centroids are applied for two parts, and k-means is applied to optimise dependence on two groups. Then one of the two groups selected, k measures are used to divide the data into three groups, initiated by dividing the centroid of the cluster into the two parts and the centroid of the remaining cluster's parts. The process must be processed until a set number of groups is reached.

Table 5. Centers of the cluster by each criterion

Clusters	Number of Items	Distinct count of OrderNumber	Count of Pieces
Cluster 1	19050	4.84	73.8
Cluster 2	4312	25.20	463.2
Cluster 3	1355	61.27	1339.0
Cluster 4	346	129.07	3136.8
Cluster 5	16	427.25	9974.9
Cluster 6	101	237.50	7160.3
Cluster 7	3	673.67	34606.0
Cluster 8	8	398.63	19198.0

The algorithm of Lloyd, which employs squared Euclidean distances, is used to compute the k-means clustering for each k is used. In combination with the splitting procedure to identify the first centres for each $k > 1$, the resulting clustering is deterministic, with the result depends only on the number of clusters.

The analysis of variance for the model is delivered. The results are presented in Table 6.

Table 6. Center of the clusters by each criterion

Variable	F-statistic	p-value	Sum of Squares	Sum of Squares
Distinct count of OrderNumber	3342.0	0.0	36.34	39.12
Count of Pieces	3029.0	0.0	12.33	14.65

The statistics for clustering is identified and presented below. One-way or one-factor ANOVA employs **F statistics** as a part of the variance explained by the variable. This is the ratio of the variance between the groups of the whole variant. The higher the F statistic, the better the relevant variable distinguishes groups.

The value of p is the probability that the F distribution of all possible values of the F statistic is larger than the actual statistic of the variable F. If the value p falls under the significance level indicated below, the null hypothesis (that the individual elements of the variable are random samples from one population) is rejected. The degrees of freedom of this F distribution is

(k - 1, N - k), where k is the number of groups (clusters), and N is the number of grouped elements (rows). The lower the value of p is the more likely the values of the elements of the respective variable differ per cluster.

The sum of the squares of the models is the ratio of squares sum between the groups and the degrees of freedom for the constructed model.

The sum of squares between groups is a measure of the change in cluster measures. If the cluster measures are close to each other (pluses are close to the overall average), this value is very small. The model has k-1 degrees of freedom, where k is the number of clusters.

The sum of the squares of the error is the ratio of squares sum within the group to the degrees of freedom of error. The sum of the squares within a cluster measures the differences in observations for each group. The error has N-k degrees of freedom, where N is the total number of observed groups (rows) and k is the number of groups (clusters). The sum of the squares of the error can be considered as the total root mean square error when the centre of each cluster is properly selected.

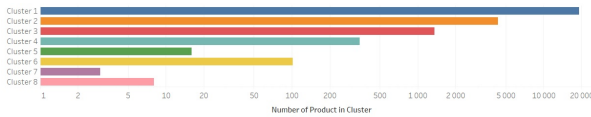


Figure 7. Numbers products in each cluster

The product is relocated to from the existing stocking place to the new one. For every location, the time of moving from the packing area is calculated. It is delivered to define the usefulness of this location and to rank it. The calculation of this time is calculated from the area located near the first gate (coordinate: x=335; y=0) by the formula:

$$Ta(i) = \frac{ABS(335 - x_i) + y_i}{1000} * \varphi * v + n * \sigma \quad (2)$$

The products are placed in the warehouse based on minimising the total time required for completing the products, described by the formula (3).

$$PO_{\min} : (t_{pp}) = \sum_{x=1}^{xa} \sum_{y=1}^{ya} \sum_{p=1}^{pa} (t_a(x, y) + t_p) \quad (3)$$

3.5 Effectiveness evaluation

The time for orders picking in the analysed period is used for results comparison. For the Reference Variant, the time for every case is known. For the variant with product relocation, the simulation of the picking process is processed.

The picking route is constructed by dividing data to separate orders and storekeepers. In Figure 8, the sequence of processed locations during the picking is presented. The place visited by different storekeeper is marked by a different colour.

The algorithm which is used for the time calculation is constructed and presented below in specified steps.

Step 1 – This step figures out the absolute value of coordinates subtraction and calculates time between actual and destination point (4). For the number of racks in a row it is used x, and for the number of rows - y.

$$x_{temp} = |x_{pocz} - x_{konc}| \quad (4)$$

$$y_{temp} = |y_{pocz} - y_{konc}|$$

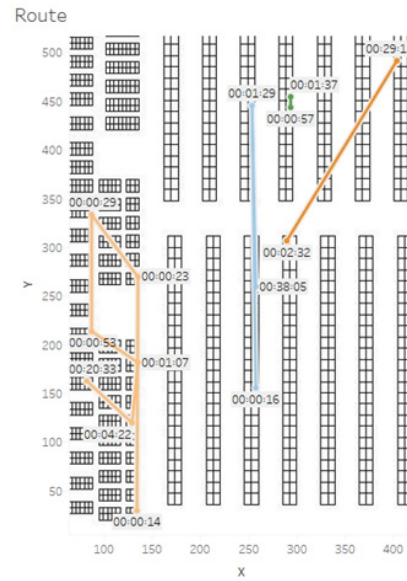


Figure 8. Sample of visited locations sequence by the storekeeper

Step 2 – During the second step algorithm checks if both racks are in the same row, and in case one row is coming out of the packing zone, it is treated as a start point. That condition is formulated as (5).

$$x_{temp} \neq 0 \text{ and } x_{temp} \neq 2 \quad (5)$$

Step 3 – When the condition described in step 3 is applied, it is necessary to include two turns into the time calculation $l_{luk} = 2$. Moreover, if the rack consists of two rows, for the product, stocked access is available; also, longitudinal aisles must be included. The exact number of that aisle is calculated by the formula (6).

$$l_{korwz} = \frac{y_{temp}}{3} \quad (6)$$

In that case, the length of the route across the warehouse is described the formula (7).

$$dr_{poprz} = D_{wr} (y_{temp} - l_{korwz}) + d_{cr} \cdot l_{korwz} \quad (7)$$

Step 4 – In case the warehouse has one transverse aisle, it is necessary to check condition if both actual and destination racks are located in the same or different warehouse zone. For such revision formula (8) is applied.

$$\left(x_{pocz} \leq \frac{l_{reg}}{2} \text{ and } x_{konc} > \frac{l_{reg}}{2} \right) \text{ or} \quad (8)$$

$$\left(x_{konc} \leq \frac{l_{reg}}{2} \text{ and } x_{pocz} > \frac{l_{reg}}{2} \right)$$

Step 5 – If the condition presented in step 4 is fulfilled, the route along the warehouse (dr_{wzd}) can be estimated by applying formula (9).

$$dr_{wzd} = D_{lr} \cdot d_{cr} \left(|x_{pocz} - x_{konc}| - 1 \right) \quad (9)$$

Step 6 – In case, the condition presented in step 4 is not matched, there is a need to check in which zone of the researched warehouse (upfront or down front the transverse aisle) both racks are located (10).

$$x_{pocz} < \frac{l_{reg}}{2} \quad (10)$$

Step 7 – If the condition described in step 6 is applicable, this means that both racks are in the first (nearest) zone of the warehouse, so if:

$$D_{lr} \cdot d_{cr}(x_{pocz} + x_{konc}) < D_{lr} \cdot d_{cr}(l_{reg} - x_{pocz} - x_{konc}) \quad (11)$$

that:

$$dr_{wzd} = D_{lr} \cdot d_{cr}(x_{pocz} + x_{konc}) \quad (12)$$

else:

$$dr_{wzd} = D_{lr} \cdot d_{cr}(l_{reg} - x_{pocz} - x_{konc}) \quad (13)$$

Step 8 - If the condition in step 6 is not fulfilled, both racks are in the second zone of the warehouse. So a new condition is checked:

$$D_{lr} \cdot d_{cr}(x_{pocz} + x_{konc} - l_{reg}) < D_{lr} \cdot d_{cr}(2l_{reg} - x_{pocz} - x_{konc}) \quad (14)$$

that:

$$dr_{wzd} = D_{lr} \cdot d_{cr}(x_{pocz} + x_{konc} - l_{reg}) \quad (15)$$

else:

$$dr_{wzd} = D_{lr} \cdot d_{cr}(2l_{reg} - x_{pocz} - x_{konc}) \quad (16)$$

Step 9 – If the condition described in step 2 is not accepted, there is no need to add turns for the calculation, so $l_{luk}=0$. In such case, the length of the route in the transverse aisle is equal to null ($dr_{poprz}=0$). So, the length of the route in the longitudinal corridor depends

on the number of racks placed are in the same zone. It revises if both shelves belong to the same area. So, if formula (17) is processed:

$$\left(x_{pocz} \leq \frac{l_{reg}}{2} \text{ and } x_{konc} > \frac{l_{reg}}{2} \right) \text{ or} \quad (17)$$

$$\left(x_{konc} \leq \frac{l_{reg}}{2} \text{ and } x_{pocz} > \frac{l_{reg}}{2} \right)$$

Such means that both racks are in different zones, so:

$$dr_{wzd} = D_{lr} \cdot d_{cr}(x_{konc} - 1) \quad (18)$$

else:

$$dr_{wzd} = D_{lr} \cdot x_{konc} \quad (19)$$

Step 10 – According to designed picking route, the time for picking is calculated (20).

$$t_{pp} = t_{lp}(dr_{wzd} + dr_{poprz}) + \sigma \cdot l_{luk} + t_p \quad (20)$$

Base on the presented formulas, for each order the picking time is calculated.

4. RESULTS AND DISCUSSION

The variants presented in the paper are compared with each other for the evaluation of total picking time. The results of the comparison divided into area P, and area T are presented in Figure 9 as the box-and-whisker chart. The points in blue and orange colour mean the total picking time for each order. The vertical line shows the lower or upper whisker. The grey rectangle presents the typical results between the lower or upper hinge, and the difference between grey colour gives the median.

The summary of statistics is presented in table 7.

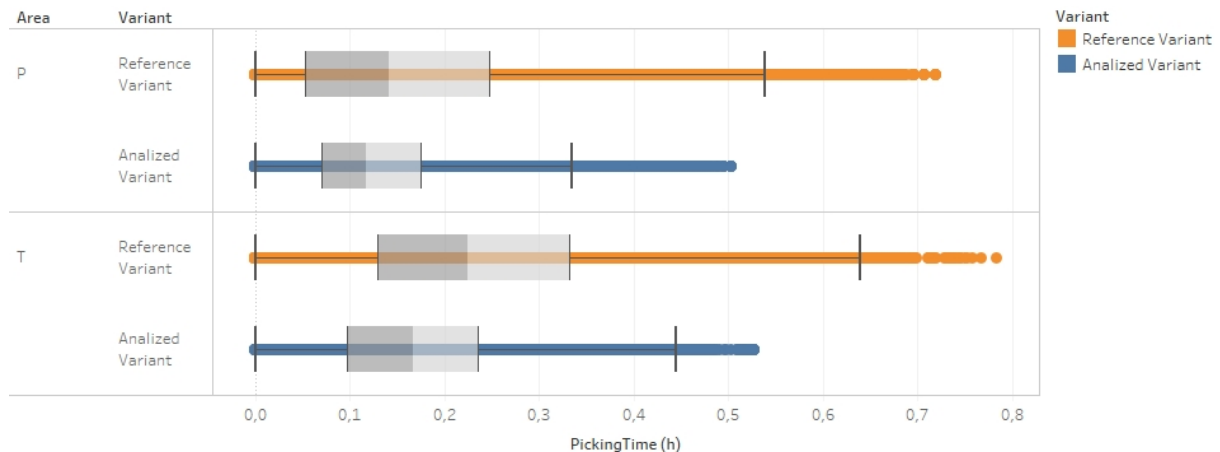


Figure 9. Sample of visited locations sequence by the storekeeper

Table 7. Centres of the cluster by each criterion

	Area P			Area T		
	Picking time [h]		The difference [%]	Picking time [h]		The difference [%]
	Reference Variant	Analysed Variant		Reference Variant	Analysed Variant	
Upper whisker	0.539	0.335	37.8 ↓	0.639	0.445	30.4 ↓
Upper hinge	0.248	0.177	28.8 ↓	0.334	0.237	29.0 ↓
Median	0.141	0.117	17.4 ↓	0.225	0.167	25.6 ↓
Lower hinge	0.054	0.071	-30.8 ↑	0.130	0.098	24.6 ↓
Lower whisker	0.004	0.004	9.8 ↓	0.003	0.002	12.0 ↓

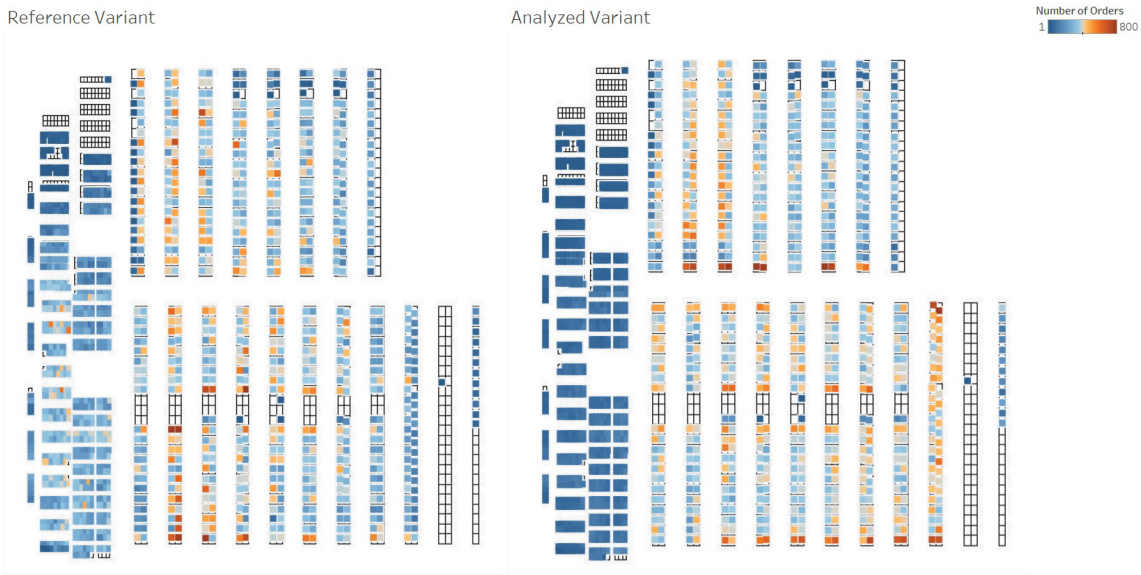


Figure 10. Sample of visited locations sequence by the storekeeper: Reference (left) and Analyzed Variant (right)

Based on the presented results, it is clear that the clustering, which is used to solve PAP, allows reducing total picking time by 17.4% ÷ 25.6%. The number depends on the type of stocking representing by the area.

The density of picking product from each location is presented in Figure 10.

5. CONCLUSIONS

The theme of BigData analysis in warehouse management emerged after the 2010s paying special attention to warehouse 4.0 and Internet of Things. During the decade, the number of papers covered the growing topic and involved other aspects of warehouse management. As a result of the literature review, the authors identified that most of the articles neither focus on order picking nor analyse e-commerce aspects.

Nevertheless, BigData is bringing a significant improvement in the field of order picking and takes an important part of the warehouse system. This improvement is linked with more efficient utilisation of resource and fulfilment of e-commerce orders. The authors delivered the case study and identified the opportunity for advancement.

The presented research shows that it is possible to use the data possessed by the company and make a simple analysis to improve the efficiency of the order picking process. In presented research row data from WMS, the layout and topology of the warehouse are used. After data cleaning, preprocessing and measuring the coordinates of each location, it is possible to solve Product Allocation Problem by relocating the products and making the simulation of order picking. The results showed that reducing the picking time is about 17.4% ÷ 25.6% when exact number depends on the area (package or pallet units).

It could be noticed that some of the locations are not demanded (marked as no coloured); these locations could be allocated for the spare space needed in the case of demand peak. In the Reference Variant, there are many red points (top of the picking scale) located

throughout the warehouse. The redistribution of products is required to ensure fast access times. Redistribution allocates products closer to the gate and the vertical warehouse corridor.

REFERENCES

- [1] Addo-Tenkorang, R. and Helo, P.T., Big data applications in operations/supply-chain management: A literature review, *Computers & Industrial Engineering*, Vol. 101, (2016) 528-543.
- [2] Arunachalam, D., Kumar, N. and Kawalek, J. P., Understanding big data analytics capabilities in supply chain management: unravelling the issues, challenges and implications for practice, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 114, (2018) 416-436.
- [3] Baker, P., Croucher, P. and Rushton, A., *The Handbook of Logistics and Distribution Management: Understanding the Supply Chain*, Kogan Page Publishers, London, (2017).
- [4] Bodden-Streubuhr, M., Warehouse-Management-Systeme im Spannungsfeld von Industrie 4.0, in Sohr, S. et al. (Ed.), *Handbuch Industrie 4.0: Produktion, Automatisierung und Logistik*, Springer, Heidelberg, (2016) 1-13.
- [5] Croxton, K.L., Lambert, D.M., García-Dastugue, S.J. and Rogers, D.S., The demand management process, *The International Journal of Logistics Management*, Vol. 13 No. 2, (2002) 51-66.
- [6] Chan, H.K., Choi, T.M. and Yue, X., Guest editorial big data analytics: Risk and operations management for industrial applications, *IEEE Transactions on Industrial Informatics*, Vol. 12 No. 3, (2016) 1214-1218.
- [7] Chongwatpol, J. and Chan, H.K., Prognostic Analysis of Defects in Manufacturing, *Industrial Management & Data Systems*, Vol. 115 No. 1, (2015) 64-87.

- [8] Choi, T.M., Wallace, S.W. and Wang, Y., Big data analytics in operations management, *Production and Operations Management*, Vol. 27 No. 10, (2018) 1868-1883.
- [9] Crisp, R.D., *How to Reduce Distribution Costs— A Practical, Scientific Approach to Increased Selling Efficiency*, Funk & Wagnalls Company, New York, (1968).
- [10] Davenport, T.H., *Analytics 3.0*, Harvard business review, Vol. 91 No. 12, (2013) 64-72.
- [11] Dremel, C., Wulf, J., Herterich, M.M., Waizmann, J.C. and Brenner, W., How AUDI AG Established Big Data Analytics in Its Digital Transformation, *MIS Quarterly Executive*, Vol. 16 No. 2, (2017) 80-89.
- [12] Egner, C.G., *Experiment in worker productivity*, working paper, University of Montana, Montana, 25 May, (1973).
- [13] Fang, M.Y. and Lin, F., Measuring the performance of ERP system—from the balanced scorecard perspectives, *Journal of American Academy of Business*, Vol. 10 No. 1, (2006) 256-263.
- [14] Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S.F., Childe, S.J., Hazen, B. and Akter, S., Big data and predictive analytics for supply chain and organisational performance, *Journal of Business Research*, Vol. 70, (2017) 308-317.
- [15] Han, Q.L. and Mei, Y.X., Distribution center: BP artificial neural network based site selection, *China Soft Sci*, Vol. 24, (2004) 140-143.
- [16] Hazen, B.T., Skipper, J.B., Boone, C.A. and Hill, R.R., Back in business: Operations research in support of big data analytics for operations and supply chain management, *Annals of Operations Research*, Vol. 270 No. 1-2, (2018) 201-211.
- [17] Hendrickson, B.A., U.S. Patent No. 5,930,784. Washington, DC, U.S. Patent and Trademark Office, (1999).
- [18] Jiang, Z.Z. and Wang, D.W., Model and algorithm of location optimisation of distribution centers for B2C E-commerce, *Control and Decision*, Vol. 20 No. 10, (2005) 11-25.
- [19] Khan, M.R., An efficiency measurement model for a computerised warehousing system, *International journal of production research*, Vol. 22 No. 3, (1984) 443-452.
- [20] Lawrence, F.B., Jennings, D.F. and Reynolds, B.E., *ERP in Distribution*, South-Western Pub, Nashville, (2005).
- [21] Lorenc, A. and Lerher, T., Effectiveness of product storage policy according to classification criteria and warehouse size, *FME Transactions*, Vol. 47 No. 1, (2019) 142–150.
- [22] Lerher, T., *Warehousing 4.0 by using shuttlebased storage and retrieval systems*, *FME Transactions*, Vol. 46 No. 3, (2018) 381-385.
- [23] Malmborg, C.J., Balachandran, S. and Kyle, D.M., A model based evaluation of a commonly used rule of thumb for warehouse layout, *Applied Mathematical Modelling*, Vol. 10 No. 2, (1986) 133-138.
- [24] Matthias, O., Fouweather, I., Gregory, I. and Vernon, A., Making sense of big data—can it transform operations management?, *International Journal of Operations & Production Management*, Vol. 37 No. 1, (2017) 37-55.
- [25] Mineo, A.M. and Plaia, A., Statistical Multivariate Techniques for the Stock Location Assignment Problem, in Rizzi, A. et al. (Ed.), *Advances in Data Science and Classification* Springer, Berlin, Heidelberg, (1998) 573-578.
- [26] Papadopoulos, T., Gunasekaran, A., Dubey, R., Altay, N., Childe, S.J. and Wamba, S.F., The role of Big Data in explaining disaster resilience in supply chains for sustainability, *Journal of Cleaner Production*, Vol. 142, (2017) 1108-1118.
- [27] Pettit, T.J., Croxton, K.L. and Fiksel, J., Ensuring supply chain resilience: development and implementation of an assessment tool, *Journal of business logistics*, Vol. 34 No. 1, (2013) 46-76.
- [28] Prasad, S., Zakaria, R. and Altay, N., Big data in humanitarian supply chain networks: A resource dependence perspective, *Annals of Operations Research*, Vol. 270 No. 1-2, (2018) 383-413.
- [29] Salah, S., Rahim, A. and Carretero, J.A., Implementation of Lean Six Sigma (LSS) in supply chain management (SCM): an integrated management philosophy, *International Journal of Transitions and Innovation Systems*, Vol. 1 No. 2, (2011) 138-162.
- [30] Singh R., *Sales and distribution management: a practice-based approach*, Vikas Publishing House, Mumbai, (2016).
- [31] Tiwari, S., Wee, H.M. and Daryanto, Y., Big data analytics in supply chain management between 2010 and 2016: Insights to industries, *Computers & Industrial Engineering*, Vol. 115, 319-330, (2018).
- [32] Waller, M.A. and Fawcett, S.E., Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management, *Journal of Business Logistics*, Vol. 34 No. 2, (2013) 77-84.
- [33] Wamba, S.F., Gunasekaran, A., Dubey, R. and Ngai, E.W., Big data analytics in operations and supply chain management, *Annals of Operations Research*, Vol. 270 No. 1-2, (2018) 1-4.
- [34] Wamba, S.F., Gunasekaran, A., Papadopoulos, T. and Ngai, E.W., Big data analytics in logistics and supply chain management, *The International Journal of Logistics Management*, Vol. 29 No. 2, (2018) 478-484.
- [35] Weberschlager, M., *Die Rolle der Global Distribution Systems (GDS) in der Online-Flugbuchung – Alternative Distributionslösungen und Mobile Commerce als Möglichkeit die Ticketdistribution zu revolutionieren*, GRIN Verlag, Munchen, (2013).

[36] Zezzatti, O. and Ochoa, C.A., Logistics Management and Optimization through Hybrid Artificial Intelligence Systems, IGI Global, Hershey, (2012).

NOMENCLATURE

SS_B – the overall between-cluster variance,
 SS_W – the overall within-cluster variation,
 k – the number of clusters,
 N – the number of observations,
 $T_a(i)$ – the time of the movement from packing area to the location i , h,
 x_i – the x coordinate of the location i , px,
 y_i – the y coordinate of the location i , py,
 φ – scale factor, for this file the $\varphi=0.118$, m/px,
 v – the speed of movement, km/h,
 n – the number of turns,
 σ – time for the penalty for the decrease of velocity and acceleration of it at each turn,
 d_{cr} – the width of aisle,
 D_{lr} – the width of racks,
 dr_{poprz} – length of the route taken across the warehouse,
 dr_{wzd} – length of the path to walk along the warehouse,
 D_{wr} – the width of the rack,
 I_c – is the correction indicator.
 I_p – the intensity work of employees, loading unit /man-hour or kg/working hour,
 K_p – the costs of pallets flow through the warehouse, EURO/quantity,
 K_{rm} – the annual costs of storage, EURO,
 l_h – the number of hours worked by the storekeeper, working hour.
 l_{korwz} – total value of longitudinal aisle, rounded up,
 l_p – the number of storekeeper, workers,
 l_{reg} – the amount of the particular rack in a row,
 l_{sp} – the number of picked order lines, quantity,
 l_{spp} – the number of items correctly picked, amount,
 p – the stocking tier $p=1, \dots, N$,
 P_k – the correctness of picking,
 P_{pmax} – is maximal value,
 P_{pmin} – is minimal value,
 P_q – product quantity,
 P_r – the annual flow of units through the warehouse, quantity,
 t_a – the additional time of penalty for forklift turn,
 t_{lp} – the time of moving 1-meter straight route,
 t_p – the time of lifting and lowering folk on specified stock level (p).
 t_{pp} – total moving time,
 t_{pr} – work time per shift, h.
 W_k – the picking productivity, quantity/working hour,

w_{om} – the turnover of warehouse understood as the sum of delivery and dispatch,
 W_p – the work productivity, loading unit/storekeeper or kg/storekeeper,
 x_{konc} – coordinate of destination rack in a row,
 x_{pocz} – coordinate of the actual rack in a row,
 x_{temp} – route length between coordinates for racks in a row,
 y_{konc} – coordinate of destination row,
 y_{pocz} – coordinate of the actual row,
 y_{temp} – the distance of the route between coordinates in a row.

ПОБОЉШАЊЕ КОМИСИОНИРАЊА НАЛОГА КОД Е-ТРОВИНЕ КОРИШЋЕЊЕМ АНАЛИЗЕ WMS И BigData ПОДАТАКА

А. Лоренц, А. Буринскиен

Истражује се побољшање процеса комисионирања налога без додатних улагања у софтвер, упосленике, алате и залихе. У циљу решавања проблема коришћени су и обрађени подаци о комисионирању из система за управљање складиштем (WMS). Анализирани су подаци из извора BigData и извршена је кластеризација производа помоћу софтвера Tableau, при чему је решаван проблем алокације производа (ПАП). Извршено је израчунавање времена комисионирања за референтни и нов сценарио и извршено је поређење. Показује се да би стандардни подаци из WMS могли да се користе за решавање ПАП проблема за скраћење укупног времена комисионирања. Метод који аутори описују може да се користи за типично складиште у коме упосленици и виљушкари обављају процес комисионирања. После побољшања план би могао да се користи за аутоматско комисионирање и примену код WMS. За анализу података BigData софтвер Tableau се повезује са базом података WMS. Дато решење би могло да се користи за свакодневну анализу података и решавање проблема алокације производа. Метод је лак за коришћење, нема нових улагања у скуп софтвер и аутоматизацију комисионирања да би се постигле велике перформансе процеса комисионирања налога. Његова примена омогућава повећање ефикасности. Власници продавница могу да бирају више производа истовремено. Истраживање је оригинално јер користи једноставне методе и анализу специфичних података који су до сада коришћени само за израчунавање показатеља перформанси упосленика.