
Optimization of Logistics Processes of the Supply Chain Using RFID Technology

Submitted 10/03/21, 1st revision 11/04/21, 2nd revision 16/05/21, accepted 30/06/21

Paweł Rymarczyk¹, Arkadiusz Małek², Ryszard Nowak³, Jacek Dziwulski⁴

Abstract:

Purpose: The aim of the article is to develop a system for optimizing supply chain logistics processes using RFID technology.

Design/Methodology/Approach: Machine learning algorithms such as Gradient Boosting, random forests, decision trees, RUS were used to solve the problem. An RFID reader and dedicated software were designed.

Findings: The results of the conducted research show that the methods used, a procect application based on RFID technology, increases the company's efficiency in logistic processes.

Practical Implications: The methods and system presented in the article can be used in the supply chain of logistics and manufacturing companies.

Originality/Value: The novelty is the appropriate selection and use of machine learning methods for the designed system that optimizes the logistics processes of returns using RFID technology. A proprietary system consisting of a dedicated reader and IT application was designed.

Keywords: Machine learning, RFID, Gradient boosting, random forests, decision trees, RUS.

JEL codes: C61, O30, E27.

Paper type: Research article.

¹Corresponding author, Netrix Group sp. z o.o. , Lublin, Poland,
e-mail: pawel.rymarczyk@netrix.com.pl

²University of Economics and Innovation in Lublin, Lublin, Poland,
e-mail: arkadiusz.malek@wsei.lublin.pl

³Same as in 2, e-mail: ryszard.nowak@wsei.lublin.pl

⁴Faculty of Management, Lublin University of Technology, Lublin, Poland,
e-mail: j.dziwulski@pollub.pl

1. Introduction

Historically, numbers and barcodes have been tracked to manage goods along the supply chain. Currently, RFID (Radio Frequency Identification) methods can monitor the condition and location of products from the moment they are produced until they reach the end customer. Being able to gain control over the management and quality of deliveries on time and anticipate the demand it enables is a game changer. At the same time, it should be noted that RFID, like any radio technology, has limitations not only related to the distance of information transmitting devices (Finkenzeller, 2010; Meng and Li, 2016), but also limitations related to the environmental conditions in which the signal is transmitted.

The use of RFID technology can significantly accelerate logistics processes in a warehouse, from the registration of goods delivered to the warehouse, through monitoring of their storage location, to navigation and registration of the movement of employees or forklifts. At the same time, the use of this technology requires each time consideration of the specific needs and specifics of the process and stored goods. The economic calculation of the installation of an RFID-based system is also considered, which, depending on the chosen solution, may turn out to be relatively high, especially in the case of active labels - the cost of label production (Wobak, Gebhart and Muehlmann 2012; Griffin and Durgin, 2009).

Moreover, RFID technology enables the identification of many objects at the same time, provided that they are within the operating range of the system (Kumar, Swanson, and Tran 2009; Wing, 2006; Jankowski-Mihułowicz and Węglarski, 2012). Currently, passive RFID tags play the most important role. First of all, they are much cheaper than active RFID tags due to the fact that they do not have batteries. As a result, they can function longer, because their life time is limited only by environmental conditions and possible mechanical damage. In addition, due to the need to have a power element, active RFID tags reach large dimensions, they are equipped with a housing. This causes their unit price to increase (Marrocco, 2007).

The aim of the research was to optimize logistics processes with the use of RFID technology in the supply chain. The analyzed company delivers its products to customers in returnable packaging. Currently, managing them is troublesome due to the number of transactions and packages. The introduction of RFID technology to manage returnable packaging will help the company track it and ensure the timely return. From the analysis carried out, we can distinguish the following problems when using returnable packaging:

- a disturbed traditional balance of cost allocation. This requires large investments in containers, additional transport costs,
- additional infrastructure for sorting empty containers,
- additional management and quality control systems,
- difficulty in managing a fleet of returnable containers,

- containers are routinely mishandled or lost, not tracked by the IT system,
- warehouse and logistics employees often do not know if the container has already returned from the customer, and if so, where exactly is it on the company's premises,
- the barcodes used to label the containers require a direct line of sight and only one of them can be read at a time.

2. Detection of Returned Orders

This section deals with detecting possible complaints (returned orders) in e-commerce product transactions using machine learning. Often times, some products ordered from websites, for example, are returned to the vendor for various reasons such as damage to goods, customer dissatisfaction, etc. This causes many disruptions to logistics and supply chain management for e-commerce. If a product is returned, with obvious shipping charges, it is then required to re-store the product in warehouses or book it in stock, etc. It is therefore beneficial for the service provider to anticipate these situations in order to be able to resolve any disruptions in advance. keep your operations running smoothly. On the basis of exemplary data, a predictive procedure (Santis, Aguiar, and Goliatt, 2017) will be presented showing the benefits of such an analysis. The data set used in the analysis contains 23 variables and 1,929,935 observations. The first five lines of the set used are shown in Figure 1.

Figure 1. Data set used

| | SKU | INV | TIM | INQ | FOR3 | FOR6 | FOR9 | SAL1 | SAL3 | SAL6 | ... | OVRP | SUP1 | SUP2 | OVRA | RSK2 | RSK3 | RSK4 | RSK5 | RSK6 | BO |
|---|---------|----------|-----|-----|------|------|------|------|------|------|-----|------|-------|----------|------|------|------|------|------|------|----|
| 0 | 1026827 | 0.000000 | 8.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.782 | 0.776896 | 0.0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 1043384 | 1.000000 | 9.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.990 | 0.990000 | 0.0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 2 | 1043696 | 1.000000 | 8.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.782 | 0.776896 | 0.0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 3 | 1043852 | 0.989949 | 8.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.100 | 0.130000 | 0.0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 4 | 1044048 | 0.872872 | 8.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... | 0.0 | 0.782 | 0.776896 | 0.0 | 1 | 0 | 0 | 1 | 0 | 0 |

Note: The individual columns mean:

SKU - Random Product ID

INV - current inventory level

TIM - lead time for the product (if available)

INQ - quantity of the product in progress

FOR3, FOR6, FOR9 - sales forecast for the next 3,6,9 months

SAL1, SAL3, SAL6, SAL9 - the number of units sold in the last month, 3,6,9 months

MIN - the minimum recommended quantity for production

RSK1 - root problem for the identified part

OVRP - overdue parts

SUP1 - source performance in the last 6 months

SUP2 - source performance in the last 12 months

OVRA - the amount of overdue orders in the warehouse

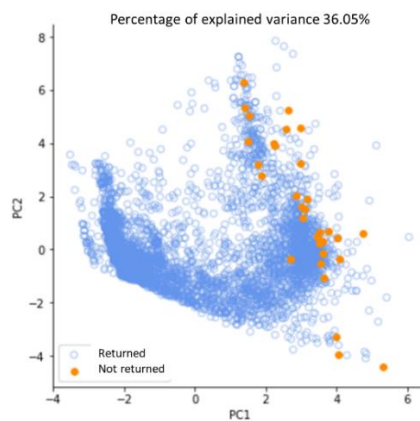
RSK2, RSK3, RSK4, RSK5, RSK6 - flags of subsequent risks

BO - purpose variable, it means whether the product has been returned or not

Source: Own creation.

It should be noted that the set presented above is already a transformed set, i.e., the relevant variables have been standardized, binarized and missing data have been handled with the help of imputation. From the data considered, a random sample of 5,000 observations was selected to illustrate the scale of the products returned. To visualize the data, the PCA analysis was used to be able to reduce the number of dimensions to two. The figure below shows the result of the above code. From the data, we can conclude that only 0.7% of all products were returned - which is a well-known problem of unsustainable data. However, thanks to the appropriate measures calculated in the next section, we are able to assess the fit to the data.

Figure 2. PCA chart of the quantity of returned products in the sample under consideration



Source: Own creation.

Decision trees are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of the target variable by learning simple decision rules inferred from the available data.

For the given training variables $x_i \in R^n, i = \{1, 2, \dots, l\}$ and the vector of the decision variable $y \in R^1$, the decision tree recursively divides the space in such a way that cases with the same labels are grouped together. If we denote the data in node m by Q , then each candidate $\theta = (j, t_m)$ consisting of the function j and the threshold t_m divides the data set into two subsets:

$$Q_{left}(\theta) = (x, y) | x_j \leq t_m, \tag{1}$$

$$Q_{right}(\theta) = Q \setminus Q_{left}(\theta). \tag{2}$$

The impurities in node m are calculated using the following function H , the choice of which depends on the problem under consideration (regression or classification)

$$G(Q, \theta) = \frac{n_{left}}{N_m} H(Q_{left}(\theta)) + \frac{n_{right}}{N_m} H(Q_{right}(\theta)). \quad (3)$$

In the problem of classification under consideration, the typical measures of pollution are the Gini index

$$H(X_m) = \sum_k p_{mk}(1 - p_{mk}), \quad (4)$$

and the entropy coefficient:

$$H(X_m) = - \sum_k p_{mk} \text{LOG}(p_{mk}), \quad (5)$$

where p_{mk} is the percentage of observations of class k at node m

$$p_{mk} = \frac{1}{N_m} \sum_{x_i \in R_m} I(y_i = k), \quad (6)$$

and X_m is part of the training data found at node m . Finally, in order to minimize the level of contamination, parameters are selected so that

$$\theta^* = \arg \text{MIN}_{\theta} G(Q, \theta), \quad (7)$$

based on recursive repetition of operations for the subsets $Q_{left}(\theta^*)$ and $Q_{right}(\theta^*)$ until the maximum allowable tree depth is reached.

Random forests is a method based on an analysis based on decision trees, creating their larger clusters, i.e. forests, and on their basis generates a new model. Its purpose is to combine the results of several basic estimators built with a given learning algorithm in order to improve the robustness of the model or its generalization. The usual method is the averaging method, in which the main principle is to build several estimators independently and then average their results. The resulting average estimate is usually better than any of the individual estimators because its variance is reduced.

Random Under Sampling (RUS) are those where there is a serious distortion in the class distribution, i.e. there is a definite difference in class size. The two main approaches to randomly re-sampling an unbalanced dataset are removing examples from the majority class called RUS under-sampling and duplicating examples from the minority class called oversampling (Random Over Sampling, ROS) (Galar *et al.*, 2012). Usually (and in our case), RUS is used as a stage before building the model. In the presented reasoning, it was used before the use of decision trees.

The Gradient Boosting Decision Trees (GBDT) method builds the model by adding additional fields and checking the percentage of explained variance. It also allows you to optimize any differentiable objective function. Based on the data (in the case under consideration, the classifying variable has two states), a regression tree is generated. Thus, mathematically, additive models of the figure are considered:

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x), \tag{8}$$

where $h_m(x)$ are basic functions, usually called weak learners in the learning process. Like the rest of the enhancement algorithms, the model is built as follows:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x), \tag{9}$$

where the newly formed tree tries to minimize the loss function taking into account the previous version:

$$h_m = \text{ARG MIN}_h \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + h(x_i)). \tag{10}$$

The initial F_0 model is determined depending on the problem. For least squares regression, we choose the mean of the target values.

The gradient boosting method tries to numerically solve the minimization problem by taking the steepest direction. The steepest direction is the negative gradient of the loss function computed in the current model F_{m-1} which can be computed for any differentiable loss function as:

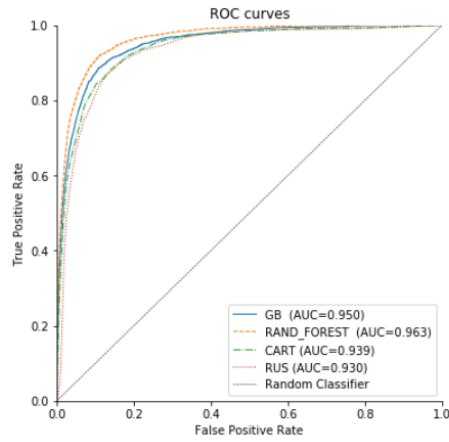
$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^n \nabla_F L(y_i, F_{m-1}(x_i)), \tag{11}$$

where the stride length is chosen as:

$$\gamma_m = \text{ARG MIN}_\gamma \sum_{i=1}^n L\left(y_i, F_{m-1}(x_i) - \gamma \frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}\right). \tag{12}$$

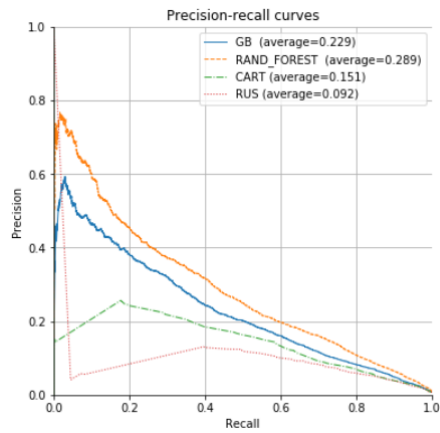
The only difference between the regression and classification algorithms is the loss function used. The code that builds the model and adjusts it to the data is presented below. The graphs obtained on the basis of the learned models are presented below.

Figure 3. ROC curves for individual models



Source: Own creation.

Figure 4. Fitting measure curves for individual models



Source: Own creation.

The quality of fit measures illustrated in the figures were calculated on the basis of the following formulas:

Precision:

$$P = \frac{T_P}{T_P + F_P}, \tag{13}$$

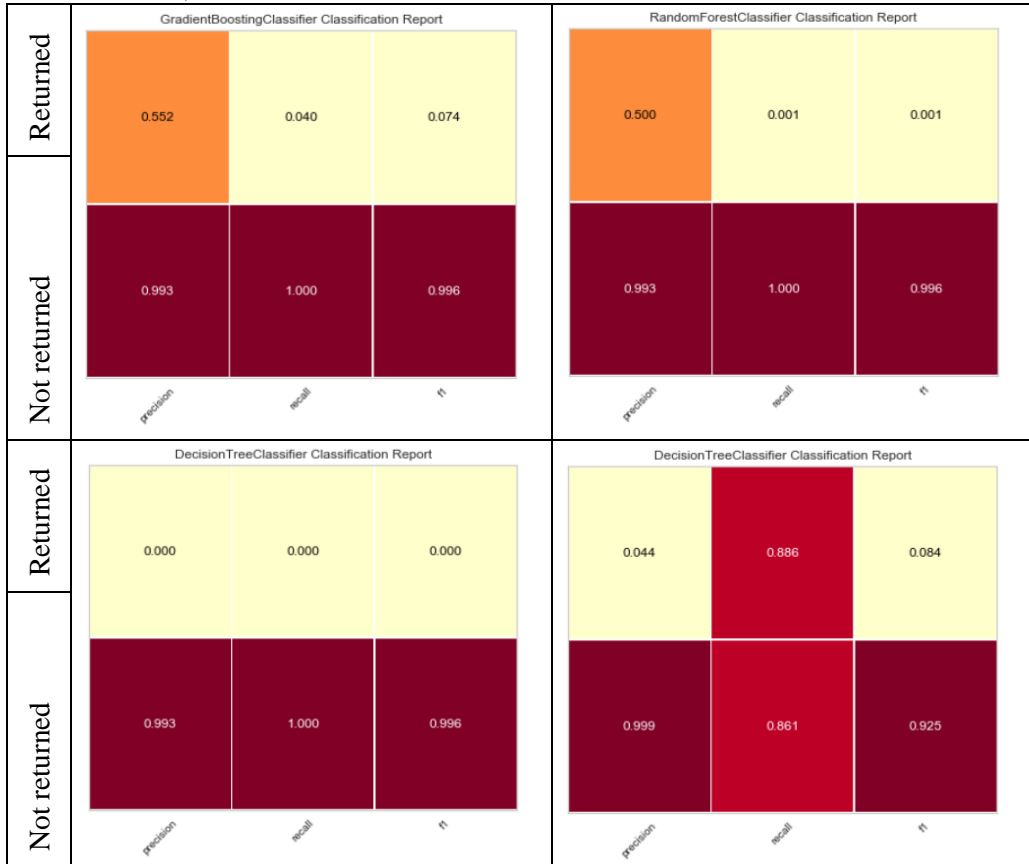
Recall:

$$R = \frac{T_P}{T_P + F_N}, \tag{14}$$

AUC:

$$AUC = \frac{1 + P - F}{2}. \tag{15}$$

Figure 5. Model quality chart using methods: Gradient Boosting, random forests, decision trees, *RUS*



Source: Own creation.

Moreover, the mean precision was calculated as:

$$AVG = \sum_n (R_n - R_{n-1})P_n, \tag{16}$$

where P_n , R_n are precision and recall respectively on the n th threshold. The following figures show the basic measures depending on the model used. The measure f_1 is also given here, calculated as:

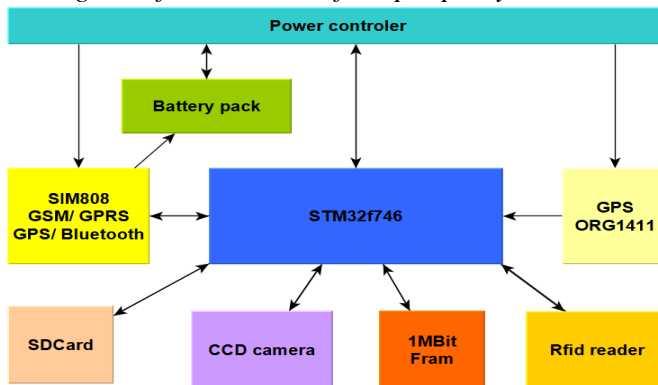
$$f_1 = 2 \frac{P \cdot R}{P + R}, \tag{17}$$

All the presented visualizations confirm that the classifier based on the theory of random forests and Gradient Boosting seem to best fit the data, but note that each of the algorithms (and to a lesser extent the RUS algorithm) loses accuracy due to the already mentioned problem of unbalanced data.

3. Project of Using RFID Technology in Logistic Processes

The main goal of the work is to design a system that allows automatic control of IBC containers in terms of the number and location of the last known location. This solution is to improve the logistics related to the resources of returnable containers and minimize the losses associated with their loss or keeping them for too long by the customer. The RFID technology was used as a source of information about the availability of a given equipment in the location (Figure 6).

Figure 6. Block diagram of connections of the periphery with the control unit



Source: Own creation.

In order to test the system, an available VAN type vehicle with a cargo space was used. The final effect of the project is the detectability of the RFID transponder, regardless of where it is located in the cargo space of the car. The object presence detection system is based on RFID technology. This technology allows to obtain information about the presence of an object marked with a tag compliant with the standard (Figure 7-9).

Figure 7. Random distribution of items with RFID tags



Source: Own creation.

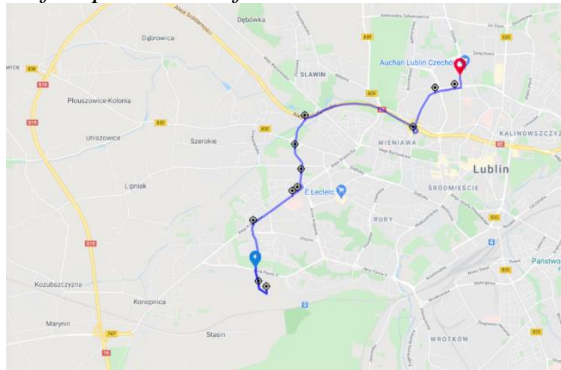
Figure 8. Presentation of the detection results of labels arranged in the same way as the elements in Figure 7

| TagID | Count | RSSI | First Seen | Last Seen |
|-------------------------|-------|------|---------------------|---------------------|
| E2034217A4D0541509855D9 | 16 | -85 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| K2030804000090587A14518 | 19 | -64 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| E20305040000409587A1119 | 17 | -59 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| E30305040000609587A1918 | 18 | -51 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| K20305040000409587A0818 | 17 | -43 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| E20305040000409587A1D18 | 16 | -36 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| F30305040000609587A0818 | 18 | -38 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| E20305040000609587A1918 | 18 | -43 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| E20305040000409587A0E18 | 17 | -27 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| K20305040000609587A1918 | 19 | -43 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| E2034217A5D0541509855EC | 16 | -43 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| E3034217A8D0541509855D4 | 18 | -49 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |
| B2034217A6B84150D86432 | 17 | -60 | 6/4/2020 9:34:58 AM | 6/4/2020 9:35:02 AM |

| Reader | Tags | Reads | Read Rate | Ant 1 | Ant 2 | Ant 3 | Ant 4 | Ant 5 | Ant 6 | Ant 7 | Ant 8 |
|------------------|------|-------|-----------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1. 192.168.0.100 | 13 | 229 | 54 | 195 | 114 | | | | | | |

Source: Own creation.

Figure 9. Location of displacement of the monitored device



Source: Own creation.

4. Conclusions

The aim of the work was to use RFID technology in the logistics processes of a production company by creating a system design that automatically allows containers to be controlled in terms of the number and location of the last known location. This solution is to improve the logistics related to the resources of returnable containers and minimize the losses associated with their loss or keeping them for too long by the customer. The created system meets these assumptions.

The designed and tested system together with the hardware infrastructure can be used for further adaptation works on the customer's vehicles and containers. The use of RFID technology in IT systems contributes to the improvement of the automatic identification of semi-finished products and products in the production process and in the further stage of their operation. The use of this technology to control products at every stage of their "life" seems very interesting and justified. Currently, the main disadvantage of implementing such solutions is primarily the cost of a single RFID

tag. However, the continuous development of the production technology of RFID tags may result in lowering their costs. Their appropriate use (e.g., multiple use of the same identifiers at the current market prices of consumables seems to be possible.

References:

- Finkenzeller, K. 2010. RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and near-Field Communication. IEEE.
- Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F. 2012. A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4), 463-484. <https://doi.org/10.1109/TSMCC.2011.2161285>.
- Griffin, J.D., Durgin, G.D. 2009. Complete link budgets for backscatter-radio and RFID systems. *IEEE Antennas and Propagation Magazine*, 51(2), 11-25. <https://doi.org/10.1109/MAP.2009.5162013>.
- Jankowski-Miśkiewicz, P., Węglarski, M. 2012. Synteza anteny czytnika / programatora indukcyjnie sprzężonego systemu RFID bliskiego zasięgu funkcjonującego w paśmie HF. *Przegląd Telekomunikacyjny i Wiadomości Telekomunikacyjne*, 8-9, 1077-1084.
- Kumar, S., Swanson, E., Tran, T. 2009. RFID in the healthcare supply chain: Usage and application. *International Journal of Health Care Quality Assurance*, 22(1), 67-81. <https://doi.org/10.1108/09526860910927961>.
- Marrocco, G. 2007. RFID antennas for the UHF remote monitoring of human subjects. *IEEE Transactions on Antennas and Propagation*, 55(6), 1862-1870. <https://doi.org/10.1109/TAP.2007.898626>.
- Meng, Z., Li, Z. 2016. RFID Tag as a Sensor - A Review on the Innovative Designs and Applications. *Measurement Science Review*, 16(6), 305-315. <https://doi.org/10.1515/msr-2016-0039>.
- Santis, R.B., Aguiar, E.P., Goliatt, L. 2017. Predicting material backorders in inventory management using machine learning. In *2017 IEEE Latin American Conference on Computational Intelligence*. IEEE. <https://doi.org/10.1109/LA-CCI.2017.8285684>.
- Wing, R. 2006. RFID applications in construction and facilities management. *Journal of Information Technology in Construction*, 11, 711-721.
- Wobak, M., Gebhart, M., Muehlmann, U. 2012. Physical limits of batteryless HF RFID transponders defined by system properties. In *2012 IEEE International Conference on RFID-Technologies and Applications*. IEEE. <https://doi.org/10.1109/RFID-TA.2012.6404500>.