

Research Article

Anna Borucka*

Logistic regression in modeling and assessment of transport services

<https://doi.org/10.1515/eng-2020-0029>

Received Aug 20, 2019; accepted Oct 16, 2019

Abstract: Transport companies operate in a dynamically developing and competitive market. Maintaining the already held position and further expansion requires adjusting the level of services provided to the needs and requirements of customers, as well as continuous surveying, monitoring and adjusting the implemented strategy. There are different methods for such an analysis. This article proposes logistic regression. The research was conducted on the basis of a distribution and trade company dealing with the supply of automotive spare parts. As the most profitable group of customers are local car repair shops, it was this group that was subject to analysis.

The quality of service assessment was considered from the point of view of delivery time. The dichotomous form of the predictor taking two values - late and on-time delivery - was determined. From among the possible ones, regressors whose influence was statistically significant and whose modification was possible were selected. The research showed which of them (and how strongly) affect the dependent variable, which allowed for modification of strategy and implementation of new solutions increasing the number of satisfied customers.

Keywords: transport, quality of service, distribution and trade company.

1 Introduction

The high quality of services provided is of key importance for maintaining a proper level of customer satisfaction and increasing the revenues of each company [1]. It is, however, a result of numerous elements, often difficult to measure, sometimes even to identify. The lack of quantifiable measures and limited availability of empirical data are also not

conducive to its analysis, which is why some research in this area are simulation studies [2, 3].

Methods aimed at studying or measuring the achieved quality of services play an important role in the process of developing high standards of customer service and its accompanying elements. They can be used to make corrections and streamline processes carried out. The most popular method in literature is SERVQUAL [4–6] or six sigma tools [7]. In many analyses, customer satisfaction is defined as a quantitative variable, most often expressed as a percentage [8, 9]. This does not always allow for a clear and immediate assessment. The possibility of prediction is also reduced due to the requirements imposed on the form of individual variables. This applies, for example, to commonly used linear regression models [10, 11]. The answer to the limitations formulated for linear models are non-linear models [12, 13]. Due to their versatility and greater freedom in the selection of variables, they are used in many sectors [14, 15], including transport [16, 17]. Therefore, they have also become a tool used in this publication, which aimed primarily to indicate the possibility to mathematically analyze selected elements of the company's activity affecting the quality of services provided. The method of estimation of such a model and its advantages are presented. The strength of the proposed approach lies in its practical usefulness and ability to provide information to shape the company's strategy. It allows to introduce appropriate corrections, influencing the increase in customer satisfaction, which is particularly important in ensuring loyalty to the company and shaping long-term relations with customers.

2 Logistic regression model

Logistic regression describes and estimates the relationship between a dependent variable Y , which takes only two possible values, resulting from the occurrence or not of an event, and independent variables affecting that phenomenon. Its popularity is supported by the lack of necessity to meet many requirements for linear regression and general linear models, which include the linearity of the re-

*Corresponding Author: Anna Borucka: Military University of Technology Warsaw, Poland; Email: anna.borucka@wat.edu.pl

relationship between a dependent and an independent variable, as well as normality and homoscedasticity of the distribution of independent variables or the necessity to use the variables given using the metric system of measures.

The logistic regression model is based on a logistic function [18, 19] that takes the form (1):

$$f(x) = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}} \quad (1)$$

where e – Euler number, x – value of the explanatory variable X .

It can be written in several ways, depending on the calculated value. If the probability of success is calculated (assuming that Y means a dichotomous variable with values of 1 – for the occurrence of the event we are interested in (success) and 0 – for the opposite case (failure), then the logistic regression model is described by the following equation (2):

$$P(Y = 1|x_1, x_2, \dots, x_k) = \frac{e^{\beta_0 + \sum_{i=1}^k \beta_i x_i}}{1 + e^{\beta_0 + \sum_{i=1}^k \beta_i x_i}} \quad (2)$$

where β_i $i = 0, \dots, k$ are logistic regression coefficients, x_1, x_2, \dots, x_k – independent variables, which can be both measurable and qualitative.

Logistic regression can also be interpreted from the point of view of the odds of occurrence of the event studied (success):

$$\frac{P(Y = 1|X)}{1 - P(Y = 1|X)} = e^{\beta_0 + \sum_{i=1}^k \beta_i x_i} \quad (3)$$

The odds are defined as the probability of an event occurring $P(A)$ divided by the probability of an event not occurring $1 - P(A)$:

$$(Odds) S(A) = \frac{P(A)}{P(non - A)} = \frac{P(A)}{1 - P(A)} \quad (4)$$

An important concept in this respect is also the odds ratio OR . It is expressed by the ratio of the odds of an event $S(A)$ occurring in one group to the same event $S(B)$ occurring in another, compared group. The OR indicator determines how much greater or lesser the odds of a given event occurring are.

$$OR_{A \times B} = \frac{S(A)}{S(B)} = \frac{P(A)}{1 - P(A)} : \frac{P(B)}{1 - P(B)} \quad (5)$$

When there is only one independent variable, the logistic regression equation takes the following form:

$$P(Y = 1|X) = \frac{e^{\beta_0 + \beta_1 \cdot x_1}}{1 + e^{\beta_0 + \beta_1 \cdot x_1}} \quad (6)$$

After logarithmizing both sides of the equation (6) we will receive a logit form of the logistic model (7):

$$\text{logit } P(Y = 1|X) = \ln \frac{P(Y = 1|X)}{1 - P(Y = 1|X)} = \beta_0 + \beta_1 \cdot x_1 \quad (7)$$

The logistic regression model has also its own requirements. Its use is conditioned by the test sample size, which should be equal to $n > 10(k + 1)$, where k means the number of predictors.

3 Characteristics of the company studied

The subject of the research is a distribution and trade company operating in the automotive industry. It deals with the sale and supply of spare parts for motor vehicles. It offers a wide range of assortment for vehicles of various brands, both passenger cars and trucks, as well as motorcycles and agricultural tractors. It also supplies additional accessories, tires, workshop tools as well as liquids and consumables.

The analysis presented here concerns one of the company's branches dealing with sales and direct delivery of orders to local car repair shops. The deliveries are carried out with the use of the company's own fleet, by drivers assigned to specific areas of the city within the area of branch operations. Orders are placed via customer service center and a special program with an individual account for accessing online catalog databases. Depending on the time of placement, orders are assembled and segregated into 4 areas serviced by 4 drivers in four rounds held each day. Timeliness of deliveries is one of the most important elements influencing the assessment of the level of services provided. It should be stressed, however, that the priority for the company is the quality of the sold assortment, and in particular its reliability and durability in relation to price. So far, therefore, the assessment of the delivery service itself, including its timeliness, has not been of particular interest to the company. The implementation arrangements once adopted in this area are consistently continued. Both the number of drivers, equal at any time of the day, and the division of the serviced area among them remains constant. In the case of repair shops, however, waiting time for parts is as important as quality thereof, since prolonged deliveries cause downtime for the mechanics and can reduce potential profits. Moreover, offers of competing companies in this industry often include products from the same manufacturers and similar discounts are offered to the part-purchasing companies. Therefore, any area that increases competitiveness is worth considering.

The above considerations became the genesis of the presented research. The research sample consisted of data on the completion of individual orders provided by the company. They contained information about the round

number, congestion on the road, the number of orders placed by a given customer, the total number of repair shops serviced within a given route. In addition, it was noted whether, in the context of the order placed, the goods were available at the warehouse of the branch or had to be delivered from the central warehouse and whether the delivery was made within the time limits declared by the company. An analysis of more than 2,300 orders revealed that as much as 37% of them were late. In so many cases (862 out of 2304) the order was not delivered to the customer within the declared time.

4 Diagnostics and characteristics of selected predictors

4.1 Ordering round

As mentioned above, the courses to customers are carried out in 4 rounds resulting from the time periods in which the customer orders are placed. The first-round handles orders placed between 6 p.m. and 9 a.m., the second one those placed between 9 a.m. and 1 p.m., the third - between 1 p.m. and 3 p.m. and the fourth - between 3 p.m. and 6 p.m. It is the company's policy that items ordered after 6 p.m. should be delivered to the customer by 11 a.m. the next working day, those ordered between 9 a.m. and 1 p.m. should be delivered no later than 3 p.m., the third round should be delivered to the customer by 5 p.m., while the last round should be delivered to the repair shops by 8 p.m.

Figure 1 This has resulted in an even distribution of the number of drivers among the deliveries. Each of them at a given time is performed by 1 driver (4 trips per shift, 16 deliveries per day).

The analysis of the significance of differences in the averages of orders in particular rounds carried out using the

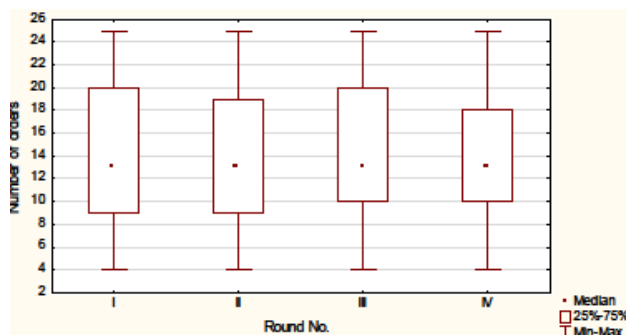


Figure 1: Comparison of the number of orders placed in particular time periods (rounds).

Kruskal-Wallis test confirmed that with the assumed level of significance ($\alpha = 0.05$) there are no grounds to reject the verified zero hypothesis ($p = 0.3384$) of the equality of averages in the groups studied. However, the analysis of late deliveries in individual hourly intervals leads to different conclusions. It shows that the most problematic group is the second one, accounting for more than 50% of all late deliveries. The results for all groups are presented in Table 1.

Table 1: Percentage share of late deliveries by individual delivery ranges.

Round No.	1st	2nd	3rd	4th
Number of late deliveries [%]	15.1	51.2	23.5	10.2

Therefore, the variable time of placing an order may be a factor affecting the quality of the services provided.

4.2 Location of the ordered assortment

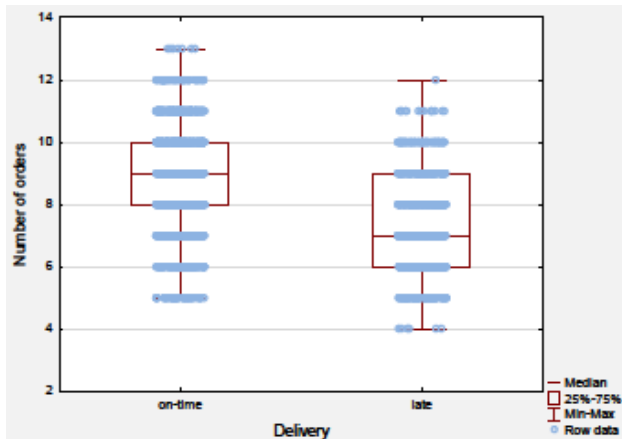
Each of the branches of the company studied has its own warehouse, where orders are assembled and prepared for delivery within each round. However, due to the diversity and a wide range of products offered, it is not possible to guarantee the availability of all items. In such a situation, the missing goods are sent from a central warehouse located in a different district of the city, which may extend the waiting time for delivery and even cause the order to be sent in the next round. As the analysis shows, the number of late deliveries is significantly higher (almost 67% of all orders) if the ordered goods are located in the central warehouse.

4.3 Congestion on the road

The next step was to characterize the congestion factor, which may have a strong impact on the delay. Congestion on roads is caused by many factors. They are influenced by both peak hours, during which there is an increase in the number of vehicles due to commuting, but also by random events such as accidents. In the period under study, delays during heavy traffic accounted for almost 64% of all late deliveries.

Table 2: Basic descriptive statistics on the number of orders placed.

Number of observations	Mean	Median	Minimum	Maximum	Standard deviation	Coefficient of variation [%]
2304	8.32	8	4	13	1.79	21.59

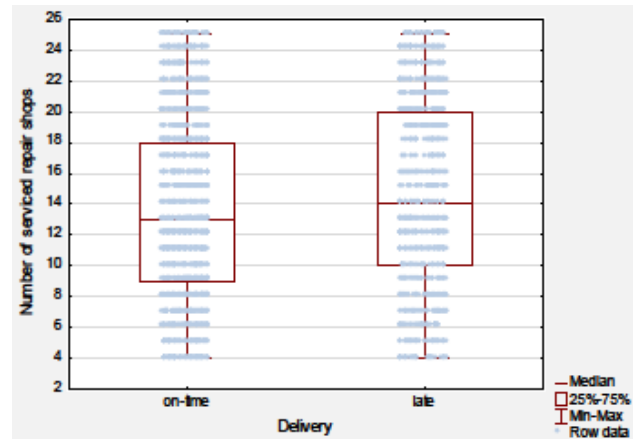
**Figure 2:** Comparison of delivery times depending on the number of orders.

4.4 Number of orders and number of serviced repair shops

Two quantitative variables were also distinguished. It was the number of orders placed by a given repair shop and the number of repair shops serviced within a given round. The number of orders is not tantamount to the number of pieces of a particular assortment. It is defined as a one-time contact (by telephone or online) with a company, within which the customer can order an unlimited number of items. Often, new orders are created on an ongoing basis as part of vehicle diagnostics, after the inspection of subsequent components or cars. During the period considered, one repair shop placed an average of 8 orders per round. Other descriptive statistics are presented in Table 2.

It appears that the number of late deliveries is lower in the case of large orders (Figure 2). This may result from their value, which in the case such orders is high, which is conducive to greater care for such a customer. A comparison of the number of late orders in these two groups is shown in Figure 2. A clearly visible difference was confirmed by the results of the Mann-Whitney U test ($p = 0.000$), which at the assumed level of significance ($\alpha = 0.05$) indicate the necessity of rejecting the verified zero hypothesis of equal averages in the group.

The analysis of the number of repair shops variable also showed its significant impact on the timeliness of delivery. It turned out that the number of late deliveries is

**Figure 3:** Comparison of delivery timeliness depending on the number of serviced repair shops.

higher when the number of repair shops serviced within a given round is higher (Figure 3).

This is confirmed by the results of the Mann-Whitney U test ($p = 0.000$), which at an assumed level of significance ($\alpha = 0.05$) indicate rejection of the zero hypothesis being verified.

Selection of predictors in the logistic regression model constitutes an important stage in the estimation of its parameters. It should be correct and exhaustive, therefore the selected and characterized above explanatory variables should be analyzed to determine whether the suggested relationships are strong enough (significant) to justify the use of the variable in the model. The chi-square test [20] allows for a statistical and substantive study of the relationship between variables. In all cases, the calculated test statistic did not allow to accept the zero hypothesis. It was therefore rejected in favor of the alternative hypothesis of the existence of a relationship, the strength of which was measured using Φ -the Yule's (for binary tables) and the V-Cramér coefficient (tables more complex than 2×2). All links between variables are statistically significant. This is also confirmed by the graphs of interaction of individual dependent variables with the explained variable (Figure 4). In the absence of interaction, the lines in the graph would be parallel to each other.

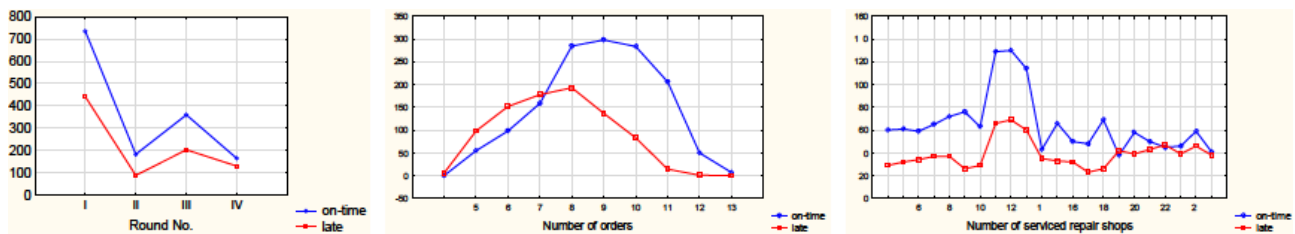


Figure 4: Some of interaction graphs of a dependent variable and predictors.

Table 3: Parameters of the logistic regression model and their assessment.

Effect	Level	Score	Standard Error	Wald's Stat.	Conf. limit upper 95.00%	Conf. limit lower 95.00%	p
number of orders		−0.297	0.019	243.004	−0.334	−0.260	0.000
number of repair shops		0.062	0.007	72.632	0.048	0.077	0.000
congestion	yes	0.579	0.094	38.280	0.396	0.762	0.000
central warehouse	no	−0.438	0.095	21.260	−0.625	−0.252	0.000
Round No.	1st	1.221	0.184	43.803	0.859	1.582	0.000
Round No.	2nd	0.850	0.147	33.545	0.562	1.138	0.000
Round No.	3rd	0.763	0.163	21.975	0.444	1.082	0.000

5 Estimation of logistic model parameters

Calculations carried out and graphs (Figure 4) confirm that the model variables were selected correctly. This allows to estimate the parameters of the logistic regression model, the values of which are presented in Table 3.

All calculated parameters turned out to be statistically significant, which is confirmed by the calculated Wald's statistics test value and the associated probability value p , which for each line is smaller than the assumed level of significance $\alpha = 0.05$. (Table 3). It can therefore be concluded that all the factors distinguished significantly contribute to the delivery being late. This is described by the logistic regression model in the following form:

$$P(\text{late delivery} = \text{yes} | X) = \frac{e^a}{1 + e^a} \quad (8)$$

where:

$$\begin{aligned} a = & -0.297 * \text{number of orders} \\ & + 0.062 * \text{number of repair shops} \\ & + 0.579 * \text{congestion} - 0.438 * \text{centr. warehouse} \\ & + 1.221 * 1\text{st round} + 0.85 * 2\text{nd round} \\ & + 0.763 * 3\text{rd round} \end{aligned} \quad (9)$$

The logistic regression equation presented above can also take equivalent form of a logit function:

$$\text{logit } P(\text{late delivery} = 1 | X) \quad (10)$$

$$= \ln \frac{P(\text{late delivery} = 1 | X)}{1 - P(\text{late delivery} = 1 | X)}$$

$$= -0.297 * \text{number of orders}$$

$$+ 0.062 * \text{number of repair shops}$$

$$+ 0.579 * \text{congestion} - 0.438 * \text{centr. warehouse}$$

$$+ 1.221 * 1\text{st round} + 0.85 * 2\text{nd round}$$

$$+ 0.763 * 3\text{rd round}$$

They can also be written as the odds:

$$P \frac{P(\text{late delivery} = 1 | X)}{1 - P(\text{late delivery} = 1 | X)} = e^a \quad (11)$$

where

$$a = -0.297 * \text{number of orders} \quad (12)$$

$$+ 0.062 * \text{number of repair shops}$$

$$+ 0.579 * \text{congestion} - 0.438 * \text{centr. warehouse}$$

$$+ 1.221 * 1\text{st round} + 0.85 * 2\text{nd round}$$

$$+ 0.763 * 3\text{rd round}$$

Estimated values of logistic regression coefficients are not a measure that quantifies the existing relationships between variables (as in the linear regression model). Only the sign at the estimated parameter informs whether the analyzed odds are greater (plus) or smaller (minus) in relation to the reference level. The information provided by the estimated parameters of the logistic regression model is supplemented by the odds calculated according to the equation (12). The odds ratio calculated on the basis

Table 4: Odds ratios for individual predictors

	Effect	Odds ratios	Upper conf. limit 95.0%	Lower conf. limit 95.0%	p
number of orders		0.743	0.716	0.771	0.000
number of repair shops		1.064	1.049	01.sie	0.000
congestion	peak	1.784	1.485	2.143	0.000
central warehouse	no	0.645	0.535	0.777	0.000
round no.	1st	3.389	2.361	4.865	0.000
round no.	2nd	lut.34	1.755	03.gru	0.000
round no.	3rd	2.144	1.559	2.949	0.000

thereof – equation (5) – indicates the size of the change signaled by the estimated parameters. The odds ratios calculated for the parameters studied are presented in Table 4.

The odds ratio calculated for the number of orders placed by a given repair shop is 0.74 and means that each unit increase in this number causes a 0.74-fold decrease in the odds of late delivery. The impact of the number of repair shops that are serviced within one round is opposite. A unit increase in this number causes a 1.06-fold increase in the odds of late delivery. Out of the qualitative variables, only the lack of the need to obtain goods from the central warehouse causes a 0.65-fold decrease in the odds of late delivery, which means that if the goods are not available in the branch warehouse, there is a 1.55-fold increase in the odds of late delivery. As the reference point was the 4th round, where the least late deliveries took place, the calculated odds ratios for the 1st, 2nd and 3rd round increase the odds of late delivery by 3.39, 2.34 and 2.14 times, respectively.

6 Logistic regression model diagnostics

The adequacy of model matching to the empirical data was verified using the Hosmer-Lemeshow test and the ROC curve [9, 21]. The Hosmer-Lemeshow test is a test that verifies the hypothesis of equality of observed and predicted values. If they are close enough to each other, it can be assumed that the model is well matched to the data. The zero hypothesis of this test is that the probabilities estimated by the model correspond to the real probabilities. The value of the H-L test statistics calculated for the analyzed model is 8.9437 and the associated p-value = 0.3471, which means that there are no grounds for rejecting H_0 .

To evaluate the proposed model, the so-called cut-off point π_0 was also used. This is the value of the predictor that best divides the studied set into two groups: the group

in which the studied phenomenon occurs and the group in which the phenomenon does not occur. This value falls within the range $[0, 1]$ and is defined as follows [22], if:

$$\hat{\pi}(x) = \hat{P}(Y = 1 | x) > \pi_0 \quad (13)$$

it is assumed that an event has occurred ($\hat{y} = 1$). In the opposite situation, when

$$\hat{\pi}(x) \leq \pi_0 \quad (14)$$

it is assumed that an event has not occurred ($\hat{y} = 0$).

The notions of sensitivity and specificity are related to the cut-off point [23]. Sensitivity is the ability of a model to detect units with a distinguished feature, determines the number of correctly predicted cases in a set of all observed occurrences [24, 25]. It is expressed by the following formula:

$$\text{Sensitivity} = SE = \frac{TP}{TP + FN} \quad (15)$$

On the other hand, accuracy determines the number of correctly predicted cases in relation to all cases.

$$\text{Accuracy} = ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

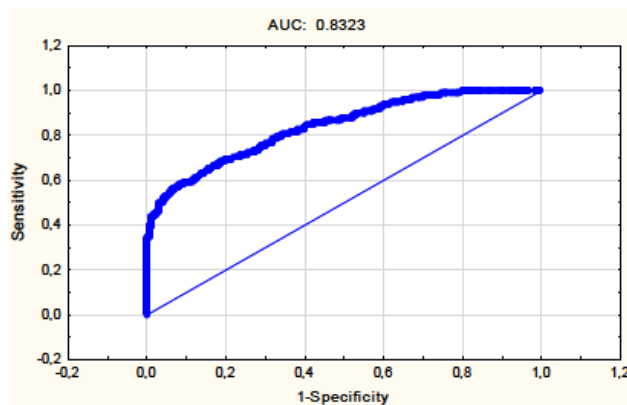
where: TP – number of true positive results, FP – number of true negative results, FP – number of false positive results, FN – number of false negative results.

The relationship between given model's sensitivity and specificity is illustrated by the ROC curve [26, 27]. It is a set of values of sensitivity and specificity calculated for each possible cut-off point, marked in the coordinate system, in which ($1 - \text{specificity}$) is represented on the axis of abscissas, while *sensitivity* on the axis of ordinates. The cut-off point closest to the point of coordinates (0,1) is called the optimum cut-off point. It is a value that optimally divides the set studied into two groups (in which the phenomenon occurs and in which it does not occur). It is determined using the Youden index (J) [22], according to the equation (17):

$$J = \text{sensitivity} + \text{specificity} - 1 \quad (17)$$

Table 5: Cut-off point of the logistic regression model studied.

Observation no.	Cut-off point	True positive	True negative	False positive	False negative	SE	1-SP	Youden index
469	0.405	498	1079	363	364	0.578	0.252	0.326

**Figure 5:** ROC curve for the logistic regression model studied.**Table 6:** AUC classification

AUC value	Score
$0.9 < \text{AUC} < 1.0$	Excellent discrimination
$0.8 < \text{AUC} < 0.9$	Good discrimination
$0.7 < \text{AUC} < 0.8$	Sufficient discrimination
$0.6 < \text{AUC} < 0.7$	Weak discrimination
$0.5 < \text{AUC} < 0.6$	Insufficient discrimination

For the proposed cut-off point, the sensitivity is 0.578, while the specificity is 0.748. There are 1,577 well classified cases (496 true positive and 1,079 true negative). There are 727 badly classified cases (363 false positive and 364 false negative cases). The ROC curve takes the form (Figure 4).

The most important parameter for assessing the ROC curve is *AUC* (*Area Under the ROC Curve*) that shows the classification quality of the analyzed diagnostic variable [28, 29]. It takes values from 0 to 1 and determines the ability of the test to distinguish between normal and abnormal results. The higher the AUC value, the better the observations are assigned to the individual groups and the better the model is. Detailed interpretation of the result based on Kleinbaum and Klein classification [30] (Table 6), shows that discrimination for the model under analysis is good (Figure 5).

The model developed can therefore be considered satisfactory. Its analysis makes it possible to assess the timeliness of deliveries from the perspective of selected predictors. They point to the need to revise the distribution strat-

egy according to the results obtained, which can significantly reduce the number of late deliveries and increase customer satisfaction.

The model shows, first of all, the significant influence of the order execution time, i.e. the round. Despite the fact that the number of orders placed is constant in individual ranges, the probability of delay is the highest for the 1st round. Therefore, it is worth considering adding an additional course during this time. Such a driver could also deal with those deliveries that are sourced from the central warehouse so that they would be delivered on time. Special attention should also be paid to small orders. It may turn out that they are placed by new customers who are looking for new suppliers. In such a situation, focusing attention only on large (expensive) deliveries may result in the loss of potential customers.

Further research should include a detailed analysis of the individual routes. Such analyses could lead to a different division of the serviced area among the drivers and, as a result, to even more favorable time conditions for deliveries.

7 Summary

The aim of the article was to indicate the possibility of mathematical assessment of selected elements of the company's activity influencing the quality of services provided and basing thereof to propose changes in shaping the company's strategy in the area of distribution. The proposed logistic regression model made it possible to analyze qualitative variables, which are often omitted from the assessment of company's activities and used only for comparative purposes or in a general study of its performance. The presented model allows for their broader analysis and inference showing the impact of individual predictors on the analyzed variable – late delivery in the case in question.

The author's intention was to present an alternative concept to most frequently undertaken in enterprises quantitative analyses and modeling with the use of linear models, in the case of which the requirements for variables are more restrictive. The advantage of this approach consists also in the unambiguous assessment of the explained

variable, which takes only two possible values of the delivery assessment – on time or not. Such an application does not give rise to a subjective interpretation (as in the case of the variable specified in %) and makes judgment much easier.

After collecting additional empirical data, the model can be extended by further variables and after introducing changes in the distribution area, it will allow to assess the effectiveness and efficiency of implemented solutions.

References

- [1] Młynarski S, Pilch R, Smolnik M, Szybka J. Methodology of network systems reliability assessment on the example of urban transport. *Eksplotacja i Niezawodność – Maintenance and Reliability*. 2018; 20(2): 278–283.
- [2] Wasiak M, Jacyna-Gółda I, Markowska K, Jachimowski R, Kłodawski M, Izdebski M. The use of a supply chain configuration model to assess the reliability of logistics processes. *Eksplotacja i Niezawodność – Maintenance and Reliability*. 2019; 21(3): 367–374.
- [3] Cunningham L, Young C, Lee M. Methodological Triangulation in Measuring Public Transportation Service Quality *Transportation Journal*. 2000; 40(1): 35–47.
- [4] Yilmaz V, Ari E, Gürbüz H. Investigating the relationship between service quality dimensions, customer satisfaction and loyalty in Turkish banking sector: An application of structural equation model. *International Journal of Bank Marketing*. 2018; 36(3): 423–440.
- [5] Awasthi A, Chauhan SS, Omrani H, Panahi A. A hybrid approach based on SERVQUAL and fuzzy TOPSIS for evaluating transportation service quality. *Computers & Industrial Engineering*. 2011; 61(3): 637–646.
- [6] Randheer K, Al-Motawa AA, Vijay PJ. Measuring Commuters' Perception on Service Quality Using SERVQUAL in Public Transportation. *International Journal of Marketing Studies*. 2011; 3(1): 21–34.
- [7] Politis Y, Grigoroudis E. Related information Using Kano's model to incorporate Six Sigma analysis in customer satisfaction measurement. *International Journal of Decision Support Systems*. 2018; 3(3–4): 146–166.
- [8] Hussain R, Al Nasser A, Hussain YK. Service quality and customer satisfaction of a UAE-based airline: An empirical investigation. *Journal of Air Transport Management*. 2015; 42: 167–175.
- [9] Kasiri AL, Guan Cheng KT, Sambasivan M, Sidin S. Integration of standardization and customization: Impact on service quality, customer satisfaction, and loyalty. *Journal of Retailing and Consumer Services*. 2017; 35: 91–97.
- [10] Elmeguid SMA, Ragheb MA, Tantawi PI, Elsamadicy AM. Customer satisfaction in sharing economy the case of ridesharing service in Alexandria, Egypt. *The Business and Management Review*. 2018; 9(4): 373–382.
- [11] Radojevic T, Stanisic N, Stanic N, Davidson R. The effects of traveling for business on customer satisfaction with hotel services. *Tourism Management*. 2018; 67: 326–341.
- [12] Hwang S, Suh EK. An Empirical Study on Nonlinear Relationship between Product Modularity and Customer Satisfaction. *Journal of Industrial Distribution & Business*. 2018; 9(2): 47–55.
- [13] Dehghan-Nayeri M, Mehregan M. Nonlinear Multi attribute Satisfaction Analysis (N-MUSA): Preference disaggregation approach to satisfaction. *Iranian Journal of Management Studies*. 2018; 11(1): 1–22.
- [14] Guillén M, Pérez-Marín AM, Alcañiz M. A logistic regression approach to estimating customer profit loss due to lapses in insurance. *Insurance Markets and Companies*. 2011; 2(2): 42–54.
- [15] Caigny AD, Coussement K, Bock KWD, A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*. 2018; 269(2): 760–772.
- [16] Bellizzi MG, Eболи L, Forciniti C, Mazzulla G. Air Transport Passengers' Satisfaction: an Ordered Logit Model. *Transportation Research Procedia*. 2018; 33: 147–154.
- [17] Świdorski A, Borucka A, Jacyna-Gółda I, Szczepański E. Wear of brake system components in various operating conditions of vehicle in the transport company. *Eksplotacja i Niezawodność – Maintenance and Reliability*. 2019; 1(29): 1–9.
- [18] Ghosh J, Li Y, Mitra R. On the Use of Cauchy Prior Distributions for Bayesian Logistic Regression. *Bayesian Anal.* 2018; 13(2): 359–383.
- [19] Wang HY, Zhu R, Ma P. Optimal Subsampling for Large Sample Logistic Regression. *Journal of the American Statistical Association*. 2018; 113:522: 829–844.
- [20] Stewart G, Kamata A, Miles R, Grandoit E, Mandelbaum F, Quinn C, et al. Predicting mental health help seeking orientations among diverse Undergraduates: An ordinal logistic regression analysis. *Journal of Affective Disorders*. 2019; 257: 271–280.
- [21] Ing E, Su W, Schonlau M, Torun N. Support Vector Machines and logistic regression to predict temporal artery biopsy outcomes, *Canadian Journal of Ophthalmology*. 2019; 54(1): 116–118.
- [22] Stanisław A. Modele regresji logistycznej. Zastosowanie w medycynie, naukach przyrodniczych i społecznych. Statsoft; 2016.
- [23] Pepe MS. The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press; 2003.
- [24] Cheng LS, Tu WJ, Shen Y, Zhang LJ, Ji K. Combination of High-Sensitivity C-Reactive Protein and Homocysteine Predicts the Post-Stroke Depression in Patients with Ischemic Stroke, *Mol Neurobiol*. 2018; 55(4): 2952–2958.
- [25] Chen F, Chen S, Ma X. Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. *Journal of Safety Research*. 2018; 65: 153–159.
- [26] Mason C, Twomey J, Wright D, Whitman L. Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression. 2018; 59(3): 382–400.
- [27] Chen WJ, Yan X, Zhao Z, Hong H, Bui DT, Pradhan B. Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBF Network models for the Long County area (China). *Bulletin of Engineering Geology and the Environment*. 2018; 78: 247–266.
- [28] Kalil AC, Mattei J, Florescu DF, Sun J, Kalil RS. Recommendations for the assessment and reporting of multivariable logistic regression in transplantation literature, *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons*. 2010; 10(7): 1686–1694.

- [29] Dou J, Yamagishi H, Zhu Z, Yunus AP, Chen CW. TXT-tool 1.081-6.1 A comparative study of the binary logistic regression (BLR) and artificial neural network (ANN) models for gis-based spatial predicting landslides at a regional scale, Landslide dynamics: ISDR-ICL landslide interactive teaching tools. Springer; 2018.
- [30] Kleinbaum DG, Klein M. Logistic Regression A Self-Learning Text. Springer; 2010.