Applied Computer Systems
ISSN 2255-8691 (online)
ISSN 2255-8683 (print)
December 2020, vol. 25, no. 2, pp. 172–179
https://doi.org/10.2478/acss-2020-0019
https://content.sciendo.com

Estimating the Efforts of Mobile Application Development in the Planning Phase Using Nonlinear Regression Analysis

Sergiy Prykhodko[1*], Natalia Prykhodko[2], Kateryna Knyrik[3]
[1–3]*Admiral Makarov National University of Shipbuilding, Mykolaiv, Ukraine*

Abstract **– The authors consider the construction of a nonlinear multiple regression model, its confidence and prediction intervals to evaluate the efforts of mobile application development in the planning phase based on the multivariate normalizing transformation and outlier detection. The constructed model is compared to the linear regression model and nonlinear regression models based on the univariate transformations, such as the decimal logarithm, Box–Cox, and Johnson transformation. This model, in comparison with other regression models, has better prediction accuracy.**

Keywords **– Estimation, maximum likelihood estimation, mobile applications, parameter estimation, planning, regression analysis, transform.**

I. Introduction

Evaluating software development efforts is one of the important problems during the planning phase for the software project manager to be able to successfully plan the software project. Like web application (app) development, mobile app development has its roots in more traditional software development [1]. However, there are differences, for example, the use of the agile methodology for mobile app development. Today, the solution to the problem of evaluating software development efforts is carried out, applying regression equations and models. One of the well-known regression equations for estimating mobile app development efforts is COCOMO II. However, its use for mobile applications has some difficulties. First, the main factor for this equation is the software size, which is still unknown in the planning phase. Second, this equation is built on a univariate transformation in the form of a decimal logarithm, which does not always allow for good normalization of the data. Third, a regression equation does not include random variables [2]–[4] and the effort estimation model based on the function point analysis method [5]. As we know, the effort is a random variable. Therefore, over the last years, several regression models for estimating the efforts of the mobile app development in a planning phase were proposed [6]–[8]. In [8], the authors built a nonlinear multiple regression model to evaluate the efforts of mobile application development in the planning phase based on the multivariate normalizing transformation and outlier detection. However, at the last iteration when constructing this model, the relative accuracy of parameter estimators was 1 %, which could affect its quality. Therefore, there is a need to improve this model, primarily according to parameter estimates.

As in [8], to improve the nonlinear regression model for estimating the efforts of developing mobile apps in the planning phase, we shall further use the method based on the multivariate normalizing transformation and outlier detection. This method consists of four steps. In the first step, a set of multivariate non-Gaussian data are normalized using a multivariate normalizing transformation. After that, normalized data are checked for multivariate outliers, and, if ones are detected, they are removed. The method based on the squared Mahalanobis distance [9] is used for outlier detection. In the second step, the nonlinear regression model is built based on the multivariate normalizing transformation. In the third step, the prediction intervals of nonlinear regression are constructed. Finally, in the fourth step, it is checked whether among the data for which the nonlinear regression model was constructed, there are data that go beyond the found bounds of the prediction interval of regression. If the outliers are detected, they are removed, and we repeat all the steps, starting with the first, for new data.

II. Construction of the Model

As in [8], we shall construct a three-factor nonlinear regression model to evaluate the effort Y (in person-hours) of developing the mobile apps in the planning phase based on the four-dimensional data set of the 38 mobile apps (see Table I). This model is built around the Requirement Analysis Document metrics of the mobile app: number of screens X_1, number of functions X_2, and number of files X_3.

The three-factor nonlinear regression model to evaluate the efforts of developing the mobile apps in the planning phase is constructed based on the Johnson four-variate transformation for S_B family according to [8] and has the form

$$Y = \hat{\varphi}_Y + \hat{\lambda}_Y \Big/ \left\{1 + \exp\left[-\left(\hat{Z}_Y + \varepsilon - \hat{\gamma}_Y\right)\Big/\hat{\eta}_Y\right]\right\}, \qquad (1)$$

[*] Corresponding author's e-mail: sergiy.prykhodko@nuos.edu.ua

where ε is a Gaussian random variable defined residuals, $\varepsilon \sim N(0,\sigma_{\varepsilon}^2)$; $\hat{Z}_Y$ is a prediction result by linear regression equation $\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3$ for normalized data, which are transformed using the Johnson four-variate transformation for S_B family with components

$$Z_j = \gamma_j + \eta_j \ln\left[\left(X_j - \varphi_j\right)/\left(\varphi_j + \lambda_j - X_j\right)\right], \qquad (2)$$

where Z_j is a standard Gaussian variable, $Z_j \sim N(0,1)$; γ_j, η_j, φ_j, and λ_j are parameters of the Johnson transformation for S_B family, $\eta_j > 0$, $\varphi_j < X_j < \varphi_j + \lambda_j$, $\lambda_j > 0$, $j = 1, 2, 3$. The component Z_Y is defined analogously (2) with the only difference that instead of Z_j, X_j, γ_j, η_j, φ_j, λ_j one should put Z_Y, Y, γ_Y, η_Y, φ_Y, λ_Y, respectively.

For the four-dimensional data set of the 38 mobile apps from Table I, the estimators of parameters for the Johnson four-variate transformation for S_B family are: $\hat{\gamma}_Y = 5.69898$, $\hat{\gamma}_1 = 0.524119$, $\hat{\gamma}_2 = 0.77618$, $\hat{\gamma}_3 = 0.540973$, $\hat{\eta}_Y = 2.40219$, $\hat{\eta}_1 = 0.74388$, $\hat{\eta}_2 = 0.79545$, $\hat{\eta}_3 = 0.534447$, $\hat{\phi}_Y = -114.545$, $\hat{\phi}_1 = 1.7242$, $\hat{\phi}_2 = 1.6885$, $\hat{\phi}_3 = 0.90$, $\hat{\lambda}_Y = 3328.564$, $\hat{\lambda}_1 = 12.3743$, $\hat{\lambda}_2 = 12.091$, $\hat{\lambda}_3 = 8.30648$; the estimators for parameters of the linear regression equation for normalized data are: $\hat{b}_0 = 0$, $\hat{b}_1 = 0.808152$, $\hat{b}_2 = -0.928296$, $\hat{b}_3 = 0.854262$. In this case, the sum of squared residuals (SSR) is 198104.0 for (1).

TABLE I

THE DATA SET, LOWER AND UPPER BOUNDS OF NONLINEAR REGRESSION PREDICTION INTERVALS BEFORE AND AFTER OUTLIER CUT-OFF

No	Y	X_1	X_2	X_3	SMD_1	LB_1	UB_1	SMD_2	LB_2	UB_2	LB_3	UB_3	LB_4	UB_4	LB_5	UB_5	SMD_6	LB_6	UB_6
1	192	5	4	3	0.4	60.5	377.3	0.3	91.9	289.1	93.9	280.4	112.4	270.3	131.4	228.5	1.0	141.6	216.6
2	272	5	4	3	1.6	60.5	377.3	3.6	91.9	289.1	93.9	280.4	**112.4**	**270.3**	-	-	-	-	-
3	288	3	2	2	12.1	88.6	524.1	11.5	153.0	362.1	179.3	367.2	180.8	345.2	218.1	332.6	10.3	233.6	323.2
4	116	6	6	4	1.0	51.1	352.9	1.5	75.7	268.2	77.4	258.6	91.6	242.6	105.7	195.7	5.3	115.6	185.7
5	372	5	5	4	5.6	**54.5**	**362.4**	-	-	-	-	-	-	-	-	-	-	-	-
6	504	9	8	6	7.5	**90.1**	**453.3**	-	-	-	-	-	-	-	-	-	-	-	-
7	28	6	7	2	5.9	−0.7	232.5	7.3	20.1	161.0	20.0	137.5	26.9	102.8	25.1	70.9	9.2	25.1	64.6
8	176	6	7	3	3.0	18.9	277.8	5.0	37.9	204.6	37.8	186.5	**45.1**	**155.6**	-	-	-	-	-
9	364	10	11	9	7.3	157.4	665.2	8.7	267.9	404.2	292.7	401.1	324.0	396.7	331.6	411.5	8.7	342.2	409.2
10	120	10	10	5	3.9	48.7	363.8	4.1	73.4	272.3	72.6	258.7	81.8	235.1	97.1	188.5	6.1	110.8	183.4
11	22	6	5	4	11.9	**70.8**	**402.2**	-	-	-	-	-	-	-	-	-	-	-	-
12	224	11	6	2	6.2	73.5	447.0	5.5	111.9	322.1	109.5	309.7	119.5	288.6	148.5	255.8	5.6	171.7	256.4
13	24	2	2	1	5.7	−23.9	170.9	6.4	8.8	123.5	12.8	109.7	16.0	59.3	15.3	51.1	6.5	15.4	49.1
14	200	11	7	4	4.0	106.5	511.6	4.9	155.9	351.0	161.9	345.7	185.2	337.3	**214.3**	**318.8**	-	-	-
15	160	6	6	7	3.8	100.6	490.0	5.9	148.3	344.0	**162.1**	**344.0**	-	-	-	-	-	-	-
16	120	2	2	1	7.6	−23.9	170.9	10.1	8.8	123.5	**12.8**	**109.7**	-	-	-	-	-	-	-
17	96	4	4	1	9.5	−33.4	149.2	11.2	**1.9**	**95.0**	-	-	-	-	-	-	-	-	-
18	202	6	5	4	0.2	70.8	402.2	0.1	103.4	301.7	107.2	295.3	128.7	287.7	148.4	248.2	0.2	160.4	237.1
19	145	4	3	2	1.3	49.2	353.3	1.5	80.8	277.2	81.5	265.9	95.2	250.3	115.3	209.6	2.0	124.7	197.8
20	198	6	5	4	0.2	70.8	402.2	0.1	103.4	301.7	107.2	295.3	128.7	287.7	148.4	248.2	0.1	160.4	237.6
21	146	4	3	2	1.3	49.2	353.3	1.4	80.8	277.2	81.5	265.9	95.2	250.3	115.3	209.6	1.9	124.7	197.8
22	191	6	6	5	0.5	66.2	392.5	0.5	96.4	294.7	101.0	289.3	122.1	281.4	139.8	238.7	0.4	151.1	227.6
23	99	3	3	2	1.6	24.7	290.0	1.6	51.0	229.4	52.1	216.3	62.3	194.1	73.0	149.9	1.3	76.4	136.2
24	382	11	12	9	7.7	140.1	624.9	9.6	257.6	400.9	275.9	396.5	311.0	393.1	317.2	402.2	8.8	326.6	397.3
25	270	9	10	8	3.2	93.4	477.2	3.2	138.4	338.8	149.0	336.6	181.3	335.4	202.6	309.0	3.3	218.5	301.1
26	282	12	7	3	7.0	104.6	532.5	7.0	163.8	362.1	169.2	356.3	190.7	346.1	223.9	332.1	6.5	246.7	331.1
27	213	10	5	2	5.1	78.5	452.7	4.3	117.2	324.4	115.4	312.5	128.8	295.5	158.6	265.3	4.4	181.0	264.8
28	322	11	7	5	3.3	126.8	560.3	3.6	184.6	367.4	195.2	363.7	226.5	359.5	257.5	355.1	2.9	278.4	354.9
29	290	10	6	4	2.7	109.1	513.1	2.5	157.5	350.7	164.2	345.6	190.7	339.5	219.5	322.5	2.4	239.9	320.2
30	223	7	7	6	0.8	78.6	425.2	0.8	112.5	312.4	119.4	308.9	144.3	303.7	164.1	266.7	0.9	177.7	257.1
31	241	5	5	6	2.1	84.9	449.3	1.6	127.2	327.4	137.7	326.4	172.8	328.2	194.4	299.7	2.0	204.6	286.3
32	87	5	5	2	1.1	17.1	267.3	1.1	37.6	200.0	36.4	179.7	43.9	149.6	49.2	111.2	2.3	53.0	103.8
33	36	3	3	1	4.8	−29.0	153.6	3.5	5.3	105.7	4.8	72.5	15.1	53.6	15.2	49.6	4.2	15.2	47.4
34	216	8	7	5	0.4	77.1	418.6	0.4	108.9	307.9	113.3	302.1	133.1	292.3	153.2	253.9	0.6	168.5	246.6
35	67	5	6	2	3.0	1.4	233.2	2.8	22.4	165.0	22.4	142.8	29.7	110.4	29.0	77.0	8.2	29.0	69.6
36	115	7	7	3	1.5	31.0	306.4	1.5	50.3	228.9	49.5	212.1	56.7	182.4	64.9	137.9	3.2	72.4	131.5
37	36	2	2	1	5.4	−23.9	170.9	5.3	8.8	123.5	12.8	109.7	16.0	59.3	15.3	51.1	6.3	15.4	49.1
38	98	3	3	2	1.6	24.7	290.0	1.6	51.0	229.4	52.1	216.3	62.3	194.1	73.0	149.9	1.4	76.4	136.2

To estimate the prediction accuracy of the nonlinear regression model (1), we used the standard metrics, such as a multiple coefficient of determination R^2, a mean magnitude of relative error MMRE, and percentage of prediction at the level of magnitude of relative error (MRE), which equalled 0.25, PRED(0.25) [10], [11]. The R^2, MMRE, and PRED(0.25) values equal respectively 0.5789, 0.4933, and 0.5263 for the nonlinear regression model (1) constructed based on the Johnson four-variate transformation for the S_B family of the data set of the 38 mobile apps from Table I. These values of the standard metrics show bad prediction results of the nonlinear regression model (1) and for the linear regression model from [8] the values of R^2, MMRE, and PRED(0.25) equal 0.5449, 0.5713, and 0.5789, respectively.

Therefore, to improve the nonlinear regression model for estimating the efforts of developing mobile apps in the planning phase, we used the method based on the multivariate normalizing transformation and outlier detection. At first, normalized data were checked for outliers by the method based on the squared Mahalanobis distance (SMD) [9]. Table I contains the values of SMD at the first, second, and sixth iterations, which are denoted as SMD_1, SMD_2, SMD_6, respectively. There are no four-variate outliers in data from Table I at all iterations for 0.005 significance level because for all data rows, the SMD values are smaller than the value of the Chi-Square distribution quantile, which equals 14.86.

Then, the nonlinear regression model was built using the Johnson four-variate transformation for the S_B family in form (1). Then, the prediction intervals of nonlinear regression are constructed by the formula [8]

$$\psi_Y^{-1}\left(\hat{Z}_Y \pm t_{\alpha/2,\nu} S_{Z_Y}\left\{1+\frac{1}{N}+\left(\mathbf{z}_X^+\right)^T\left[\left(\mathbf{Z}_X^+\right)^T\mathbf{Z}_X^+\right]^{-1}\left(\mathbf{z}_X^+\right)\right\}^{1/2}\right), \qquad (3)$$

where ψ_Y is the first component of a vector of normalizing transformation, $\boldsymbol{\psi}=\{\psi_Y,\psi_1,\psi_2,\ldots,\psi_k\}^T$; k is a number of regressors or independent variables (in our case, k equals 3); $t_{\alpha/2,\nu}$ is a student's t-distribution quantile with $\alpha/2$ significance level and ν degrees of freedom; $\mathbf{z}_X^+$ is a vector with components $Z_{1_i}-\overline{Z}_1$, $Z_{2_i}-\overline{Z}_2$, …, $Z_{k_i}-\overline{Z}_k$ for i-row; $\mathbf{Z}_X^+$ is a matrix of centred regressors that contains the values of normalized data $Z_{1_i}-\overline{Z}_1$, $Z_{2_i}-\overline{Z}_2$, …, $Z_{k_i}-\overline{Z}_k$; $\overline{Z}_j=\frac{1}{N}\sum_{i=1}^{N}Z_{j_i}$, $j=1,2,\ldots,k$; $S_{Z_Y}^2=\frac{1}{\nu}\sum_{i=1}^{N}\left(Z_{Y_i}-\hat{Z}_{Y_i}\right)^2$, $\nu=N-k-1$; $\left(\mathbf{Z}_X^+\right)^T\mathbf{Z}_X^+$ is the $k\times k$ matrix

$$\left(\mathbf{Z}_X^+\right)^T\mathbf{Z}_X^+=\begin{pmatrix} S_{Z_1Z_1} & S_{Z_1Z_2} & \ldots & S_{Z_1Z_k}\\ S_{Z_1Z_2} & S_{Z_2Z_2} & \ldots & S_{Z_2Z_k}\\ \ldots & \ldots & \ldots & \ldots\\ S_{Z_1Z_k} & S_{Z_2Z_k} & \ldots & S_{Z_kZ_k}\end{pmatrix}, \qquad (4)$$

where $S_{Z_qZ_r}=\sum_{i=1}^{N}\left[Z_{q_i}-\overline{Z}_q\right]\left[Z_{r_i}-\overline{Z}_r\right]$, $q,r=1,2,\ldots,k$.

For example, if in (3) ψ_Y is the decimal logarithm function (Log10) then for our case $Z_Y=\mathrm{Log10}(Y)$, $\psi_Y^{-1}=10^{Z_Y}$; $Z_j=\mathrm{Log10}(X_j)$, $j=1,2,3$. If in (3) ψ_Y is the Johnson transformation (2) then $\psi_Y^{-1}=\varphi_Y+\lambda_Y\big/\left(1+e^{-(Z_Y-\gamma_Y)/\eta_Y}\right)$.

In the first iteration, for the data normalized by the Johnson four-variate transformation for S_B family from 38 mobile apps, the matrix (4) is the following:

$$\left(\mathbf{Z}_X^+\right)^T\mathbf{Z}_X^+=\begin{pmatrix} 38.00 & 33.14 & 26.63\\ 33.14 & 38.00 & 31.03\\ 26.63 & 31.03 & 38.00\end{pmatrix}.$$

As in [8], for the nonlinear regression model (1) with the parameter estimators obtained from the data for the 38 mobile apps, it turned out that Y values for the three mobile apps (5, 6, and 11) were outside the prediction interval calculated by (3). In Table I, the lower and upper bounds of the prediction interval obtained in the first iteration are denoted as LB_1, and UB_1, respectively. In the second iteration, data from three apps (5, 6, and 11) were removed, and data from the remaining 35 apps were used for model construction. In Table I, the row numbers with the outliers in data are highlighted in bold at the relevant iteration. A dash (-) depicts the exception of the corresponding numbers of data at the relevant iteration. For model (1) with the parameter estimators obtained from the data for the 35 mobile apps (data rows), it turned out that the value of Y for app 17 went beyond the prediction interval. There were five such iterations, after which 29 mobile apps remained (1, 3, 4, 7, 9, 10, 12, 13, 18–38). At the sixth iteration, there were no outliers; the repeat of the stages was completed, the nonlinear regression model (1) was constructed using data from 29 apps. In Table I, the lower and upper bounds of the prediction interval obtained in the sixth iteration are denoted as LB_6, and UB_6, respectively.

In the sixth iteration, for the data set of the 29 mobile apps from Table I the estimators of parameters for the Johnson four-variate transformation for S_B family are: $\hat{\gamma}_Y=0.638164$, $\hat{\gamma}_1=0.387413$, $\hat{\gamma}_2=0.84038$, $\hat{\gamma}_3=0.477514$, $\hat{\eta}_Y=1{,}12311$, $\hat{\eta}_1=0.659463$, $\hat{\eta}_2=0.8313$, $\hat{\eta}_3=0.632614$, $\hat{\phi}_Y=-28.4433$, $\hat{\phi}_1=1.82645$, $\hat{\phi}_2=1.5912$, $\hat{\phi}_3=0.657479$, $\hat{\lambda}_Y=543.1612$, $\hat{\lambda}_1=11.5548$, $\hat{\lambda}_2=12.9939$, $\hat{\lambda}_3=8.63368$; the parameter estimators of the linear regression equation for normalized data are: $\hat{b}_0=0$, $\hat{b}_1=1.17702$, $\hat{b}_2=-1.43269$, $\hat{b}_3=1.18398$.

In the sixth iteration, for the data normalized by the Johnson four-variate transformation for S_B family from 29 mobile apps, the matrix (4) is the following:

$$\left(\mathbf{Z}_X^+\right)^T \mathbf{Z}_X^+ = \begin{pmatrix} 29.00 & 24.85 & 18.65 \\ 24.85 & 29.00 & 23.62 \\ 18.65 & 23.62 & 29.00 \end{pmatrix}.$$

In the sixth iteration, the value of SSR equals 4669.6 for (1), which is 42 times less than the corresponding sum in the first iteration. The R^2, MMRE, and PRED(0.25) values equal respectively to 0.984, 0.103, and 0.862 for (1). These values of the standard metrics show good prediction results of the nonlinear regression model (1) constructed on the basis of the Johnson four-variate transformation for the S_B family of the data set of the 29 mobile apps from Table 1.

Note, in (1), the independent variables X_1, X_2, and X_3 should be changed in the ranges from 2 to 12, from 2 to 12, and from 1 to 9, respectively.

III. Comparison of Regression Models

For comparison of the model (1) with other models, we constructed a linear regression model and nonlinear regression models based on the univariate decimal logarithm transformation (Log10), the Box–Cox transformation, and the Johnson univariate transformation for the S_B family for data of the 29 apps from Table I. The three-factor linear regression model based on the data from 29 apps has the form

$$Y = \hat{Y} + \varepsilon = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \hat{b}_3 X_3 + \varepsilon, \qquad (5)$$

where $\hat{b}_0 = 36.8484$, $\hat{b}_1 = 31.3019$, $\hat{b}_2 = -43.4205$, $\hat{b}_3 = 50.5983$.

The value of SSR equals 40908.4 for (5), which is almost 9 times the corresponding amount for the model (1). The R^2, MMRE, and PRED(0.25) values equal 0.857, 0.228, and 0.793 respectively for (5). Although these values indicate good quality of the model (5) with estimates of the parameters obtained from 29 mobile apps (according to Table I), they are worse than the corresponding values for model (1) based on the Johnson four-variate transformation for S_B family.

The nonlinear regression model is built based on the Log10 transformation for data of 29 apps from Table I in the form

$$Y = 10^{\varepsilon + \hat{b}_0} X_1^{\hat{b}_1} X_2^{\hat{b}_2} X_3^{\hat{b}_3}, \qquad (6)$$

where $\hat{b}_0 = 1.73898$, $\hat{b}_1 = 1.6687$, $\hat{b}_2 = -2.1116$, $\hat{b}_3 = 1.30125$.

The value of SSR equals 53624.7 for (6), which is more than 11 times the corresponding amount for the model (1). The R^2, MMRE, and PRED(0.25) values equal 0.812, 0.198, and 0.690 respectively for (6). These values of the standard metrics are worse than the corresponding values for the model (1) based on the Johnson four-variate transformation for the S_B family. Besides, the value of PRED(0.25) for (6) indicates the low percentage of prediction of the model (6).

The nonlinear regression model is built based on the Box–Cox univariate transformation for data of 29 apps in the form

$$Y = \left[\hat{\lambda}_Y \left(\hat{Z}_Y + \varepsilon\right) + 1\right]^{1/\hat{\lambda}_Y}, \qquad (7)$$

where $\hat{Z}_Y$ is a prediction result by linear regression equation $\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2 + \hat{b}_3 Z_3$ for normalized data, which are transformed using the Box–Cox univariate transformation [12]

$$Z_j = x(\lambda_j) = \begin{cases} \left(X_j^{\lambda_j} - 1\right)/\lambda_j, & \text{if} \quad \lambda_j \neq 0; \\ \ln(X_j), & \text{if} \quad \lambda_j = 0. \end{cases} \qquad (8)$$

Here $j = 1, 2, 3$. The variable Z_Y is defined analogously (8) with the only difference that instead of Z_j, X_j, and λ_j there should be Z_Y, Y, and λ_Y, respectively.

The parameter of transformation (8) for each variable is estimated by the maximum likelihood method [13]

$$\hat{\lambda} = \arg\max_{\lambda} l(X, \lambda), \qquad (9)$$

where the log-likelihood function is

$$l(X, \lambda) = C - \frac{N}{2} \ln \sum_{i=1}^{N} \frac{\left[x_i(\lambda) - \bar{x}(\lambda)\right]^2}{N} + (\lambda - 1) \sum_{i=1}^{N} \ln(x_i).$$

Here C is a constant, which is determined from the normalization condition; $\bar{x}(\lambda) = \sum_{i=1}^{N} x_i(\lambda)/N$; $x_i(\lambda)$ is the i-value of $x(\lambda_j)$ or Z_j from (8).

For the data set of the 29 mobile apps from Table I, the estimators for parameters of the Box–Cox univariate transformations for each of variables Y, X_1, X_2, and X_3 using the maximum likelihood method (9) are $\hat{\lambda}_Y = 390.1038$, $\hat{\lambda}_1 = 12.1000$, $\hat{\lambda}_2 = 12.4637$, and $\hat{\lambda}_3 = 9.5277$, respectively. The parameter estimators of the linear regression equation for normalized data by the Box–Cox univariate transformation are as: $\hat{b}_0 = 21.5734$, $\hat{b}_1 = 16.0849$, $\hat{b}_2 = -26.0436$, $\hat{b}_3 = 30.7021$. Parameters of the linear regression equation for normalized data were estimated by the least square method.

The value of SSR equals 49743.2 for (7), which is more than 10 times the corresponding amount for the model (1). The R^2, MMRE, and PRED(0.25) values equal 0.826, 0.180, and 0.690 respectively for (7). These values of the standard metrics are worse than the corresponding values for the model (1) based on the Johnson four-variate transformation for the S_B family. Besides, the value of PRED(0.25) for (7) indicates the low percentage of prediction of this model (7).

The three-factor nonlinear regression model is built based on the Johnson univariate transformation for the S_B family for data of the 29 apps from Table I. This model is analogous to (1) with the only difference that the data for each variable are normalized by the Johnson univariate transformation for the S_B family using the maximum likelihood method [14]

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} l(X, \boldsymbol{\theta}), \tag{10}$$

where $\boldsymbol{\theta} = \{\gamma, \eta, \varphi, \lambda\}$ is the parameter vector, the log-likelihood function is

$$l(X, \boldsymbol{\theta}) = N \ln(\eta\lambda) - \frac{N \ln(2\pi)}{2} - \sum_{i=1}^{N} \ln(x_i - \phi) - \sum_{i=1}^{N} \ln(\phi + \lambda - x_i) - \frac{1}{2} \sum_{i=1}^{N} \left[\gamma + \eta \ln \frac{x_i - \phi}{\phi + \lambda - x_i} \right]^2 .$$

Here x_i is the *i*-value of X_j or Y from (2).

For the data set of the 29 mobile apps from Table I the estimators for parameters of the Johnson univariate transformations for S_B family for each of variables Y, X_1, X_2, and X_3 using the maximum likelihood method (10) are: $\hat{\gamma}_Y = 0.250621$, $\hat{\gamma}_1 = 0.147151$, $\hat{\gamma}_2 = 0.471451$, $\hat{\gamma}_3 = 0.605927$, $\hat{\eta}_Y = 0.548155$, $\hat{\eta}_1 = 0.519404$, $\hat{\eta}_2 = 0.558891$, $\hat{\eta}_3 = 0.575457$, $\hat{\varphi}_Y = 21.1791$, $\hat{\varphi}_1 = 1.90$, $\hat{\varphi}_2 = 1.90$, $\hat{\varphi}_3 = 0.90$, $\hat{\lambda}_Y = 367.078$, $\hat{\lambda}_1 = 10.2064$, $\hat{\lambda}_2 = 10.3804$, $\hat{\lambda}_3 = 8.58177$. The parameter estimators of the linear regression equation for normalized data by the Johnson univariate transformation for S_B family are: $\hat{b}_0 = 0$, $\hat{b}_1 = 1.19747$, $\hat{b}_2 = -1.43924$, $\hat{b}_3 = 1.22071$.

The value of SSR equals 39265.6 for (1) with the estimators for parameters of the Johnson univariate transformations for the S_B family. This value of SSR is more than 8 times the corresponding amount for (1) with the estimators for parameters of the Johnson multivariate transformations for the S_B family. The R^2, MMRE, and PRED(0.25) values equal 0.863, 0.188, and 0.690 respectively for (1) with the estimators for parameters of the Johnson univariate transformations for S_B family. These values of the standard metrics are worse than the corresponding values for the model (1) based on the Johnson four-variate transformation for the S_B family. Besides, the value of PRED(0.25) for (1) with the estimators for parameters of the Johnson univariate transformations for the S_B family indicates the low percentage of prediction of this model. Note, SSR, R^2, MMRE, and PRED(0.25) values equal 9978.1, 0.965, 0.117, and 0.867 respectively for (1) with the estimators for parameters of the Johnson four-variate transformations for S_B family from [8] for 30 mobile apps. This value of SSR is more than 2 times the corresponding amount for (1) with the estimators for parameters of the Johnson multivariate transformations for the S_B family in the case of 29 apps. The values of R^2 and MMRE are worse than the corresponding values for the model (1) based on the Johnson four-variate transformation for the S_B family in the case of 29 apps. The values of PRED(0.25) are the same in both cases for 29 and 30 mobile apps. The model (1) based on the Johnson four-variate transformation for the S_B family has smaller widths of confidence and prediction intervals compared to other models, including the one from [8]. The prediction intervals of nonlinear regressions are defined by (3). The confidence intervals of nonlinear regressions are calculated the same as (3) with the only difference that in the sum in curly brackets, there will not be 1. Prediction results $\hat{Y}$, lower (LB), and upper (UB) bounds of confidence intervals for the regressions are shown in Table II. The LB and UB values in Table II indicate the width of the nonlinear regression confidence interval based on the Johnson four-variate transformation, which is smaller than after the Johnson univariate transformation for 25 from 29 data rows (except four with numbers 13, 26, 33, and 37), smaller than after the Box–Cox and decimal log transformations and smaller compared to the linear regression confidence interval width for all 29 data rows. Besides, two LB values for the linear regression confidence interval are negative. All LB values for the nonlinear regression confidence intervals are positive. The bounds of the linear regression confidence interval are symmetrical about the regression line. The bounds of the nonlinear regression confidence intervals are nonsymmetrical about the corresponding regression lines.

We obtained almost the same results for the prediction intervals of regressions. Lower (LB) and upper (UB) bounds of prediction intervals for the regressions are shown in Table III. The LB and UB values in Table III indicate the width of the nonlinear regression prediction interval based on the Johnson four-variate transformation, which is smaller than after the Johnson univariate transformation for 28 out of 29 data rows (except one with number 33), smaller than after the Box–Cox and decimal log transformation and smaller compared to the linear regression prediction interval width for all 29 data rows. Besides, six LB values for the linear regression prediction interval are negative. All LB values for the nonlinear regression prediction intervals are positive. The bounds of the linear regression prediction interval are symmetrical about the regression line. The bounds of the nonlinear regression prediction intervals are nonsymmetrical about the corresponding regression lines.

Better prediction results for the model (1) constructed by the Johnson four-variate transformation for S_B family might be explained by the best multivariate normalization of the non-Gaussian data set, which was used to build the model (1) based on this multivariate transformation. Multivariate normality was tested by SMD [15]. A condition of multivariate normality is only performed for the normalized data based on the Log10, and the Johnson four-variate transformation since for all 29 rows of the normalized data the SMD values are smaller than the value of the Chi-Square distribution quantile, which equals 11.14 for 0.025 significance level. The measures of multivariate skewness β_1 and kurtosis β_2 [16] allow one to test two hypotheses that are compatible with the assumption of multivariate normality. In our case, for 29 apps $\beta_1 = 4.14$ and $\beta_2 = 24$. The multivariate skewness and kurtosis estimators equal 9.89, 5.67, 6.23, 13.95, 6.45, and 28.01, 23.18, 23.83, 34.82, 25.09 for the data of 29 mobile apps from Table I, the normalized data based on Log10, the Box–Cox, the Johnson univariate transformations and the Johnson four-variate transformation for S_B family respectively. These estimator values indicate that the necessary but not sufficient condition for multivariate normality is approximately performed for the normalized data based on the Log10, the Box–Cox univariate transformations, and the Johnson four-variate transformation for the S_B family.

Note, we can apply the model (1) to estimate the efforts Y (in person-hours) of mobile app development in the planning phase of the discovery stage depending on the number of screens X_1, the number of functions X_2, and the number of files X_3, which should be changed in the ranges from 2 to 12, from 2 to 12 and from 1 to 9, respectively. If in (1) we substitute zero instead of the random variable ε, then using (1) we can estimate the sample mean of efforts. We can estimate the prediction intervals of the dependent random variable Y by (3). The confidence intervals of the sample mean of efforts could be defined using (3) with the only difference that in the sum in curly brackets, there will not be 1.

TABLE II

PREDICTION RESULTS, LOWER AND UPPER BOUNDS OF CONFIDENCE INTERVALS FOR REGRESSIONS

No	Y	X_1	X_2	X_3	Linear regression			Univariate transformations: Log10			Univariate transformations: Box–Cox			Univariate transformations: Johnson			Johnson four-variate transformation		
					$\hat{Y}$	LB	UB	$\hat{Y}$	LB	UB	$\hat{Y}$	LB	UB	$\hat{Y}$	LB	UB	$\hat{Y}$	LB	UB
1	192	5	4	3	171.5	151.8	191.2	182.6	161.1	206.9	191.4	169.7	214.0	168.8	130.3	210.9	177.8	170.1	185.6
3	288	3	2	2	145.1	115.7	174.5	198.0	152.6	256.9	176.3	139.7	215.7	341.3	244.3	376.4	279.4	255.4	302.7
4	116	6	6	4	166.5	149.4	183.7	151.5	135.1	169.9	178.1	159.1	197.8	152.9	116.7	194.0	148.8	141.6	156.3
7	28	6	7	2	21.9	−21.7	65.5	44.1	34.2	56.8	42.6	19.1	71.5	40.7	29.6	64.4	42.6	34.7	51.2
9	364	10	11	9	327.6	286.8	368.4	285.5	232.5	350.7	273.0	233.8	314.1	353.0	301.3	375.4	378.4	361.7	393.9
10	120	10	10	5	168.7	130.6	206.7	163.7	137.8	194.6	190.9	160.2	223.3	162.4	106.2	228.4	145.1	132.4	158.4
12	224	11	6	2	221.8	180.3	263.4	178.5	136.9	232.7	203.0	158.4	251.2	252.7	165.1	321.8	213.3	195.5	231.4
13	24	2	2	1	63.2	34.5	92.0	39.5	31.6	49.5	34.5	16.4	56.6	26.9	23.2	37.2	30.1	23.3	37.7
18	202	6	5	4	210.0	190.6	229.3	225.0	198.8	254.7	227.1	204.0	251.0	209.0	166.3	251.2	198.1	190.6	205.7
19	145	4	3	2	133.0	110.6	155.4	136.1	117.1	158.1	144.2	122.2	167.2	128.5	93.4	171.8	159.6	150.2	169.3
20	198	6	5	4	210.0	190.6	229.3	225.0	198.8	254.7	227.1	204.0	251.0	209.0	166.3	251.2	198.1	190.6	205.7
21	146	4	3	2	133.0	110.6	155.4	136.1	117.1	158.1	144.2	122.2	167.2	128.5	93.4	171.8	159.6	150.2	169.3
22	191	6	6	5	217.1	195.8	238.4	202.4	175.9	232.9	219.4	194.0	245.7	208.0	160.1	255.4	188.2	180.0	196.6
23	99	3	3	2	101.7	77.6	125.8	82.2	70.2	96.2	103.1	82.4	125.3	75.5	53.7	108.2	104.0	96.3	112.0
24	382	11	12	9	315.5	272.8	358.2	279.4	226.7	344.4	269.7	228.9	312.7	268.4	159.9	342.4	364.5	347.4	380.5
25	270	9	10	8	289.1	254.1	324.2	250.5	206.5	303.8	252.3	216.7	289.6	293.9	229.4	338.5	260.2	245.7	274.7
26	282	12	7	3	260.3	218.7	302.0	251.8	201.5	314.8	261.9	217.9	308.5	379.8	351.3	386.4	290.0	270.7	308.9
27	213	10	5	2	234.0	195.3	272.7	224.2	172.2	292.0	225.4	180.8	273.2	232.0	160.1	296.3	222.4	206.2	238.8
28	322	11	7	5	330.2	295.3	365.1	419.6	348.4	505.2	340.5	299.9	382.7	354.9	322.8	372.1	318.2	304.9	331.2
29	290	10	6	4	291.7	259.5	324.0	371.2	307.1	448.8	311.5	273.0	351.7	315.6	272.1	345.6	280.9	268.2	293.3
30	223	7	7	6	255.6	230.8	280.4	240.5	206.5	280.2	245.7	217.6	274.9	250.9	198.2	296.6	216.9	207.3	226.4
31	241	5	5	6	279.8	240.9	318.7	276.6	223.9	341.8	267.2	225.8	310.7	272.6	204.8	324.1	245.5	233.0	258.0
32	87	5	5	2	77.5	52.8	102.1	66.5	57.0	77.5	79.0	60.6	98.9	58.8	42.8	84.4	76.1	69.2	83.3
33	36	3	3	1	51.1	24.2	78.0	33.4	26.8	41.6	21.5	6.4	41.7	23.7	21.9	29.3	29.3	23.8	35.4
34	216	8	7	5	236.3	217.9	254.7	240.0	212.3	271.3	243.8	221.5	266.8	237.6	194.0	277.6	206.8	198.6	215.0
35	67	5	6	2	34.0	−1.8	69.8	44.8	35.7	56.0	46.0	24.8	71.2	42.3	31.3	63.6	47.0	39.8	54.9
36	115	7	7	3	103.8	75.0	132.6	97.8	84.0	113.9	122.0	100.5	144.9	92.7	64.9	132.1	99.6	91.0	108.6
37	36	2	2	1	63.2	34.5	92.0	39.5	31.6	49.5	34.5	16.4	56.6	26.9	23.2	37.2	30.1	23.3	37.7
38	98	3	3	2	101.7	77.6	125.8	82.2	70.2	96.2	103.1	82.4	125.3	75.5	53.7	108.2	104.0	96.3	112.0

TABLE III

LB AND UB OF PREDICTION INTERVALS FOR REGRESSION MODELS

No	Linear regression		Univariate transformations: Log10		Univariate transformations: Box–Cox		Univariate transformations: Johnson		Johnson four-variate		No	Linear regression		Univariate transformations: Log10		Univariate transformations: Box–Cox		Univariate transformations: Johnson		Johnson four-variate	
	LB	UB	LB	UB	LB	UB	LB	UB	LB	UB		LB	UB	LB	UB	LB	UB	LB	UB	LB	UB
1	85.9	257.1	107.8	309.2	108.8	288.2	44.4	340.6	141.6	216.6	25	198.8	379.5	144.9	432.9	157.2	361.3	98.2	376.9	218.5	301.1
3	56.8	233.5	111.5	351.7	92.0	277.1	136.9	384.7	233.6	323.2	26	167.2	353.5	144.1	440.2	162.5	375.8	290.2	387.7	246.7	331.1
4	81.5	251.6	89.7	256.0	98.3	272.3	40.7	332.4	115.6	185.7	27	142.1	325.8	126.0	398.9	130.6	336.0	61.6	364.9	181.0	264.8
7	−72.1	115.9	24.9	78.1	0.8	113.8	23.0	162.5	25.1	64.6	28	239.9	420.5	243.4	723.3	233.5	460.1	197.5	384.9	278.4	354.9
9	234.9	420.4	164.5	495.7	174.0	385.7	181.8	385.1	342.2	409.2	29	202.4	381.1	215.0	640.8	208.3	427.6	124.4	379.7	239.9	320.2
10	77.1	260.2	95.4	281.0	106.1	290.7	41.5	340.4	110.8	183.4	30	168.7	342.5	141.0	410.4	153.6	351.1	72.7	368.0	177.7	257.1
12	128.7	314.9	100.3	317.6	111.4	311.3	68.2	370.7	171.7	256.4	31	187.9	371.8	159.0	481.3	168.2	380.3	82.6	373.3	204.6	286.3
13	−24.9	151.3	22.6	69.1	-	99.8	21.7	79.7	15.4	49.1	32	−9.4	164.3	38.9	113.4	22.1	155.7	25.2	219.4	53.0	103.8
18	124.4	295.5	132.9	381.0	138.7	328.8	56.0	356.2	160.4	237.1	33	−36.5	138.6	19.1	58.3	-	81.0	21.4	50.7	15.2	47.4
19	46.7	219.3	79.8	232.0	70.0	234.2	35.6	317.2	124.7	197.8	34	151.0	321.6	141.8	406.2	153.3	347.3	67.3	364.3	168.5	246.6
20	124.4	295.5	132.9	381.0	138.7	328.8	56.0	356.2	160.4	237.6	35	−56.6	124.7	25.6	78.3	2.5	116.4	23.2	166.8	29.0	69.6
21	46.7	219.3	79.8	232.0	70.0	234.2	35.6	317.2	124.7	197.8	36	15.7	192.0	57.4	166.9	52.7	208.2	29.7	282.7	72.4	131.5
22	131.1	303.1	119.0	344.2	131.6	320.8	55.3	356.3	151.1	227.6	37	−24.9	151.3	22.6	69.1	-	99.8	21.7	79.7	15.4	49.1
23	15.0	188.4	48.1	140.4	38.6	185.6	27.3	256.1	76.4	136.2	38	15.0	188.4	48.1	140.4	38.6	185.6	27.3	256.1	76.4	136.2
24	221.9	409.1	160.7	485.7	170.6	382.9	71.6	375.0	326.6	397.3											

Note, we can apply other more simple regression models, for example, the model (6). But in this case, as a rule, the accuracy of the estimates will be lower. In particular, this is indicated by the widths of the confidence intervals of the regressions in Table III, which determine the accuracy of the estimate of the sample mean of efforts.

We can also use the model (1) to estimate the efforts of the development of mobile apps with a development timeline of up to three months. These apps include basic (simple) apps in which there is no back-end or a network connection, data-driven apps, and partly, authentication apps (apps with login functionality) and social networking apps.

IV. Conclusion

We have improved the three-factor nonlinear regression model for evaluating the efforts of developing mobile apps in the planning phase based on the Johnson four-variate transformation for the S_B family. This model, in comparison with other regression models (both linear and nonlinear), has a greater multiple coefficient of determination, a smaller value of the mean magnitude of relative error, a greater percentage of prediction, and smaller widths of the confidence and prediction intervals of the nonlinear regression. To construct nonlinear multiple regression models for evaluating the efforts of developing mobile apps in the planning phase, it is necessary to use multivariate normalizing transformations and outlier detection by prediction intervals. Prospects for further research may include the application of other data sets to build the multiple nonlinear regression models for estimating the efforts of developing mobile apps of a certain type that are created with a specific framework, for example, React Native.

References

[1] M. Rouse, "Mobile application development," March 2019. [Online]. Available: http://searchmicroservices.techtarget.com/definition/mobile-application-development [Accessed: Nov. 19, 2020].

[2] T. P. Ryan, *Modern regression methods*, 2nd ed. New York: John Wiley & Sons, 2008. https://doi.org/10.1002/9780470382806

[3] S. Chatterjee and J. S. Simonoff, *Handbook of regression analysis*. New York: John Wiley & Sons, 2013. https://doi.org/10.1002/9781118532843

[4] N. R. Drapper and H. Smith, *Applied regression analysis*, 3rd ed. New York: John Wiley & Sons, 1998. https://doi.org/10.1002/9781118625590

[5] T. Arnuphaptrairong and W. Suksawasd, "An empirical validation of mobile application effort estimation models," in *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS 2017)*, 2017, pp. 697–701.

[6] R. Francese, C. Gravino, M. Risi, G. Scanniello and G. Tortora, "On the use of requirements measures to predict software project and product measures in the context of Android mobile apps: A preliminary study," in *Proceedings of the 41st Euromicro Conference on Software Engineering and Advanced Applications (SEAA 2015)*, 2015, pp. 357–364. https://doi.org/10.1109/SEAA.2015.22

[7] S. A. Shahwaiz, A. A. Malik and N. Sabahat, "A parametric effort estimation model for mobile apps," in *Proceedings of the 19th International Multi-Topic Conference (INMIC 2016)*, 2016, pp. 1–6. https://doi.org/10.1109/INMIC.2016.7840114

[8] S. Prykhodko, N. Prykhodko, K. Knyrik and A. Pukhalevych, "Mathematical modeling of effort of mobile application development in a planning phase," in *Proceedings of the 1st International Workshop on Information-Communication Technologies & Embedded Systems. CEUR Workshop Proceedings*, vol. 2516, 2019, pp. 96–105.

[9] S. Prykhodko, N. Prykhodko, L. Makarova and A. Pukhalevych, "Application of the squared Mahalanobis distance for detecting outliers in multivariate non-Gaussian data," in *Proceedings of the 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET)*, 2018, pp. 962–965. https://doi.org/10.1109/TCSET.2018.8336353

[10] T. Foss, E. Stensrud, B. Kitchenham and I. Myrtveit, "A simulation study of the model evaluation criterion MMRE," *IEEE Transactions on software engineering*, vol. 11, no. 29, pp. 985–995, 2003. https://doi.org/10.1109/TSE.2003.1245300

[11] D. Port and M. Korte, "Comparative studies of the model evaluation criterions MMRE and PRED in software cost estimation research," in *Proceedings of the 2nd ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*, 2008, pp. 51–60. https://doi.org/10.1145/1414004.1414015

[12] G. E. P. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 2, no. 26, pp. 211–252, 1964. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

[13] R. A. Johnson and D. W. Wichern, *Applied multivariate statistical analysis*. 6th ed. Pearson Prentice Hall, 2007.

[14] N. Prykhodko, S. Prykhodko and M. Vorona, "The non-linear regression model to estimate the part of NPLS in the whole loan portfolio of Ukrainian banks," in *2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, 2018, pp. 261–265. https://doi.org/10.1109/SAIC.2018.8516899

[15] I. Olkin and A. R. Sampson, "Multivariate Analysis: Overview," in Smelser, N. J., Baltes, P. B. (eds.) *International Encyclopedia of Social & Behavioral Sciences*. Pergamon, 2001, pp. 10240–10247. https://doi.org/10.1016/B0-08-043076-7/00472-1

[16] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Biometrika*, vol. 57, no. 3, pp. 519–530, 1970. https://doi.org/10.1093/biomet/57.3.519

Sergiy Prykhodko is Professor and Head of the Department of Software of Automated Systems at the Institute of Computer Science and Project Management, Admiral Makarov National University of Shipbuilding. He holds a Doctor of Technical Sciences degree in mathematical modelling and computational methods from V. M. Glushkov Institute of Cybernetics of NAS of Ukraine. His research interests include mathematical modelling and simulation of random variables and stochastic processes, identification of nonlinear stochastic systems, multivariate statistical analysis, stochastic models in empirical software engineering, and computer science. He has 210 publications in the area. Contact information: Institute of Computer Science and Project Management, Admiral Makarov National University of Shipbuilding, Heroes of Ukraine Ave., 9, Mykolaiv, 54025, Ukraine.
Phone: +38-0512-424470
E-mail: sergiy.prykhodko@nuos.edu.ua
ORCID iD: https://orcid.org/0000-0002-2325-018X

Natalia Prykhodko is an Assistant Professor of the Finance Department at the Faculty of Marine Economics, Admiral Makarov National University of Shipbuilding. She holds a Doctor degree in economics, planning, and organization of management of the national economy and its sectors (industry) from Saint Petersburg Academy of Engineering and Economics (now Saint Petersburg State University of Economics) and a Doctor degree in entrepreneurship, management, and marketing from Kharkiv Sovereign State Economic University (now Kharkiv National University of Economics). Her research interests include mathematical modelling and simulation of random variables and stochastic processes, multivariate statistical analysis, stochastic models in empirical software engineering and finance. She has 48 publications in the areas. Contact information: Faculty of Marine Economics, Admiral Makarov National University of Shipbuilding, Heroes of Ukraine Ave., 9, Mykolaiv, 54025, Ukraine.
Phone: +38-063-303-71-65
E-mail: natalia.prykhodko@nuos.edu.ua
ORCID iD: https://orcid.org/0000-0002-3554-7183

Kateryna Knyrik is a Doctoral student of the Department of Software of Automated Systems at the Institute of Computer Science and Project Management, Admiral Makarov National University of Shipbuilding. She obtained Bachelor and Master degrees in computer science from Admiral Makarov National University of Shipbuilding. She has two certificates: Oracle Certified Associate Java SE 7 Programmer, and Oracle Certified Professional Java SE 7 Programmer. She has work experience in the positions of Junior, Middle, Senior, and Team Lead Android Developer. Her research interests include mathematical modelling and simulation of random variables, stochastic models in empirical software engineering, and computer science. She has 9 publications in the areas. Contact information: Institute of Computer Science and Project Management, Admiral Makarov National University of Shipbuilding, Heroes of Ukraine Ave., 9, Mykolaiv, 54025, Ukraine.
Phone: +38-0512-424470
E-mail: katrin010692@gmail.com
ORCID iD: https://orcid.org/0000-0001-8434-4035