

O. L. SOLOVEI, Candidate of Technical Sciences (Ph.D.), Kyiv National University of Construction and Architecture, Associate Professor at the Department of Information Technology for Design and Applied Mathematics, Kyiv, Ukraine, e-mail: solovey.ol@knuba.edu.ua; ORCID: <https://orcid.org/0000-0001-8774-7243>

D. A. GOBOV, Candidate of Technical Sciences (Ph.D.), National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Associate Professor at the Department of Computer Science and Software Engineering of the Faculty of Informatics and Computer Science, Kyiv, Ukraine, e-mail: d.gobov@kpi.ua; ORCID: <https://orcid.org/0000-0001-9964-0339>

PARSIMONIOUS MACHINE LEARNING MODELS IN REQUIREMENTS ELICITATION TECHNIQUES SELECTION

The subject of research in the article is machine learning algorithms used for requirement elicitation technique selection. The goal of the work is to build effective parsimonious machine learning models to predict the using particular elicitation techniques in IT projects that allow using as few predictor variables as possible without a significant deterioration in the prediction quality. The following tasks are solved in the article: design an algorithm to build parsimonious machine learning candidate models for requirement elicitation technique selection based on gathered information on practitioners' experience, assess parsimonious machine learning model accuracy, and design an algorithm for the best candidate model selection. The following methods are used: algorithm theory, statistics theory, sampling techniques, data modeling theory, and science experiments. The following results were obtained: 1) parsimonious machine learning candidate models were built for the requirement elicitation technique selection. They included less number of features that helps in the future to avoid overfitting problems associated with the best-fit models; 2) according to the proposed algorithm for best candidate selection – a single parsimonious model with satisfied performance was chosen. Conclusion: An algorithm is proposed to build parsimonious candidate models for requirement elicitation technique selection that avoids the overfitting problem. The algorithm for the best candidate model selection identifies when a parsimonious model's performance is degraded and decides on the suitable model's selection. Both proposed algorithms were successfully tested with four datasets and can be proposed for their extensions to others.

Keywords: requirements elicitation techniques, Bayesian Information Criterion, Bayes factor grades, log-likelihood, parsimonious model.

Introduction. Business analysis as an extension of requirements engineering is crucial to software development. The main business analysis deliverables are requirements and designs used as a basis for solution implementation, testing, and deployment. In turn, the critical input for the tasks of analysis, specification, and modeling of requirements and design for software is the information collected during the elicitation. Standard approaches to the requirements-gathering process have been systematized and described in the form of dozens of standard elicitation techniques. Industrial guidelines and empirical studies contain detailed descriptions of the techniques' elements and usage considerations but do not provide an elicitation selection process [1].

Consequently, one of the challenges for business analysts/requirement engineers, especially novice ones, is the selection of the appropriate requirements elicitation techniques that best fit their project. As a result, some of them are misused, others are never used, and only a few are constantly applied. To solve the problem, a machine learning model to predict/recommend using the following elicitation techniques as Interviews, Document Analysis, Process Analysis, and Interface Analysis depending on the project's context was proposed [1].

In the study [2], the model's prediction accuracy was increased by transforming the dataset from imbalanced to balanced, thus making a Random Forest Classifier learner unbiased to the majority class. Feature importance score was identified by mutual information criteria, i.e., independent from the machine learner classifier. It served as an assurance that the feature's score doesn't depend on the learning algorithm's bias. Ten features with the most

significant importance score were reported in tables 4–5 as predictors for choosing the elicitation technique.

However, in both [1] and [2], selecting the best model from the candidates remained based on the performance metrics such as Accuracy and AUC.

A model selected that way is also called a "best-fit" model. The "best fit" model is complex – it includes many parameters in order to better approximate training data. The more variables included in a model, the more dependent the model becomes on the observed data so that it can fail on the test data due to noisy, uninformative, and unrepresentative data being included in the model. i.e., a "best-fit" model is prone to overfit data [3].

Although the "best-fit" models included twenty features, we took ten features with the most significant importance score, which potentially may be incorrect if the model due to include less than ten features.

To eliminate the mentioned problems for the model proposed in works [1–2] in the current study, we will develop a parsimonious model that still accurately predicts/recommends using the techniques: Interviews, Document Analysis, Process Analysis, and Interface Analysis.

Analysis of last achievements and publications. The principle of parsimony suggests a model should be as simple as possible concerning the included variables, model structure, and several parameters. It is a desired characteristic of a model defined by a suitable trade-off between squared bias and variance of parameter estimators [4]. The construction of the parsimonious model happens in the following steps:

© Solovoei O. L., Gobov D. A., 2023



Research Article: This article was published by the publishing house of NTU "KhPI" in the collection "Bulletin of the National Technical University "KhPI" Series: System analysis, management and information technologies." This article is distributed under a Creative Common [Creative Common Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Conflict of Interest:** The author/s declared no conflict of interest.



- build multiple candidate models from the same dataset but include a different number of features;
- compare and select the best candidate as a final parsimonious model;
- assess the fit of the selected candidate.

So to develop a parsimonious model, it is required to choose the methods:

- for selecting the variables for the model;
- for comparison of the candidate models;
- for assessing the fit.
- There are three commonly used methods for selecting the variables [5–7]:
- purposeful selection;
- stepwise selection;
- best subsets.

A purposeful selection – initially, a multivariable model is built with variables which are having a significant univariate test. Then, variables skipped on the 1st step are added one by one to the model to identify the variables that are not significant by themselves but contribute to the presence of other variables. The process of deleting, refitting, and verifying continues until it appears that all of the essential variables are included in the model, and those excluded are statistically unimportant. The major problem of purposeful selection is that a final model may "overfit".

A stepwise selection is based on a statistical algorithm that checks for the "importance" of variables and either includes or excludes them based on a fixed decision rule. The "importance" of a variable is defined in terms of a measure of the statistical significance of the coefficient(s) for the variable. The statistic used for linear regression is an F -test; for logistic regression – likelihood ratio, score, and Wald test.

A "best subsets" are the number of models containing one, two, three variables, and so on, which are fitted to determine the "best" according to specified criteria.

Due to meeting the current research's goals, only the "best subsets" approach from the listed able can be applied. The statistical measure that is commonly used to compare models with different numbers of parameters based on the parsimonious principle is the Akaike Information Criterion (AIC). It measures the distance between a candidate model and the accurate model – the closer the distance, the more similar the candidate is to the truth model. AIC calculates the distance between models as expected relative to Kullback – Leibler (K–L) divergence. Although AIC is a consistent estimator of K–L divergence, there is no statistical test to compare values of AIC [8–9].

Another criterion to compare candidate models is Bayesian Information Criterion (BIC), derived from Bayesian statistical analysis and estimates. BIC approximates a Bayes factor with desirable properties for hypothesis testing and model selection [10–13]. BIC is calculated for each candidate model by equation (1)

$$\text{BIC}_d = -2 \log L_d + p \log n, \quad (1)$$

where L_d is a maximized log-likelihood; p – the number of estimable parameters included in the model and n – the number of observations in the dataset; $d = \overline{1, D}$ where D

is the number of candidate models. To calculate BIC of binary classification models with predictors matrix $X \in R^{n \times p}$ and target $y \in R^n$ which takes values in the set $\{0, 1\}$ and which is built with learner algorithms: logistic regression, support vector machine (SVC), or decision tree classifiers (RandomForestClassifier), a maximized log-likelihood from (1) is calculated as a logistic loss function:

$$L_d(y, \hat{p}) = \sum_{i=1}^n -(y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)), \quad (2)$$

where \hat{p}_i is a probability with which a fit model predicts a positive class

$$\hat{p}_i = \frac{1}{1 + \exp(-X_{ij} \bar{\beta}_j - \beta_{j_0})}, \quad (3)$$

where $i = 1, \dots, n$, $j = 1, \dots, p$, $\bar{\beta}_j$, β_{j_0} – coefficients of the model.

The final look of equation (1) is specified in (4)

$$\text{BIC}_d = -2 \log L_d(y, \hat{p}) + p \log n \quad (4)$$

The model with the smallest BIC is preferable because the complex models are almost always likely to fit the data better, so the first term in definition (4) will have a low value; however, the second provides a way to penalize these extra parameters, therefore causes BIC is increasing. Candidate models' comparison by BIC_d is done by BIC weights w_d calculated by equation

$$w_d = \frac{\exp\left(-\frac{1}{2} \Delta_d\right)}{\sum_{r=1}^R \exp\left(-\frac{1}{2} \Delta_d\right)}, \quad (5)$$

where Δ_d is a difference between d -th candidate model with calculated BIC_d and the minimum value of BIC. BIC weights are the probability that d -th candidate model is the best among the candidate models. BIC weights use Bayes factor grades to evaluate a candidate model according to the rules:

- If $w_d \leq 0.5$, then this is a weak candidate model;
- If $0.5 < w_d < 0.75$, then this is a positive candidate model;
- If $0.75 \leq w_d < 0.95$, then this is a strong candidate model;
- If $0.95 \leq w_d < 1$, then this is very strong.

To assess the goodness of fit of the selected candidate models compared to etalon (or best-fit) models in works [14] is proposed to apply testing of the hypothesis based on a difference between sample means of the model's performance metric. When the mean accuracy of the selected parsimonious models is \bar{A}_1 and the mean accuracy of best-fit models is \bar{A}_2 then the parsimonious models fit if the null hypothesis is not rejected by the computed two-tailed p -value of the t -statistic (eq. 6).

$$t = \frac{\bar{A}_1 - \bar{A}_2}{\sqrt{\frac{1}{n} \sum_{i=1}^n (A_{1i} - A_{2i})^2 - \frac{\left(\sum_{i=1}^n (A_{1i} - A_{2i}) \right)^2}{n}}}, \quad (6)$$

where n is the number of the parsimonious models included in the test; $ddof$ is the delta degree of freedom with a value equal to 1. Other classification metrics, such as AUC, f1 score, precision, recall, and Jaccard score, can be used to measure the goodness of the parsimonious model in the same manner as specified in equation 6 for the accuracy metric.

The problem statement. We aim to build parsimonious models for four datasets considered in works [1–2] to avoid overfitting problems associated with the best-fit models. To design an algorithm for assessing a parsimonious model's performance compared to the best-fit model and selecting the best candidate. To execute tests to prove that the proposed algorithms can be used with other datasets.

Experiment Methodology. Our experiment methodology for constructing and assessing the parsimonious model is specified per each phase of the supervised learner model's creation lifecycle [15].

Data preparation and acquisition. Original data was formed based on a survey conducted among business analysts and requirement engineers in Ukraine regarding their use of requirement elicitation techniques and their context. Three hundred twenty-eight practitioners completed the survey. Four respondents were disqualified due to incorrect data: non-filled industrial sector and non-filled team types. The features included in the dataset used in this study are two types:

- features to describe the project's context;
- features to list all elicitation techniques used in the project.

The following features belong to the first type:

- country;
- project size: small – till 15 people, mid-size – 15–30 people, big – 30–100 people, and very big – more than 100 people;
- industrial sector;
- company business model: IT company+outsourcing, IT company+outstaffing, IT company+product, Non-IT company;
- company size: small – till 80 employees, mid-size – 81–200 employees, upper-mid size – 201–800

employees, big – 801–1500; very big – over 1500 employees;

- Software type (current project): business, embedded, scientific, system, other;
- Team types: collocated, distributed;
- Experience in business analysis/requirement engineering: up to 3 years; 3–5 years; 5–10 years; over ten years;
- Methodology: Agile, Hybrid, plan-driven (e.g., Waterfall)
- Project Category: development from scratch, user interface engineering, solution reengineering (re-design and reimplementation), solution customization;
- business analysis activities in which the respondent is usually involved in: Business analysis planning & monitoring, Strategy analysis, Elicitation & Collaboration, Requirements analysis and design definition, Solution evaluation, and Requirements life cycle management;
- certification.

The following features belong to the second type:

- benchmarking and market analysis;
- brainstorming;
- business rules analysis;
- collaborative games;
- data mining;
- design thinking / lean startup;
- document analysis;
- interface analysis;
- interviews;
- observations;
- process analysis/process modeling;
- prototyping;
- reuse database and guidelines;
- stakeholders list, map, or personas;
- survey or questionnaire;
- workshops and focus groups.

The dataset contains information about the features, along with the names of target classes such as "Elicitation", "Document Analysis", "Interface Analysis", and "Process Analysis". However, a feature with the same name as a target class is not included in the list of features. Databases' characteristics and imbalanced ratios calculated as majority-to-minority samples are specified in table 1.

Data preprocessing. The imbalance predictors matrix X and a target vector y were transformed into balanced X^* , y^* by applying SMOTE method. This method allows us to

Table 1 – The characteristics of datasets

Target class name	Majority class	Minority class	Imbalance ratio	Machine learning task	Feature type	Missing values, Y/N?
Interviews	282	41	6.9	Binary classification	Discrete	N
Document Analysis	276	47	5.9			
Interface Analysis	232	91	2.5			
Process Analysis	213	110	1.9			

generate synthetic samples that diminish the class imbalance problem. SMOTE gets a new sample from two samples from the minority class (x^R and x) with $0 \leq k \leq 1$; x^R is randomly chosen among the 5-minority class nearest neighbors of x .

Modeling. The steps undertaken are described as per pseudocode (fig. 1).

```

Input:  $X^*$ ,  $y^*$ , Acc_min, AUC_min,  $F$ 
Output:  $\vec{S}$ 
 $\vec{S} = []$ 
1. FOR  $i$  in range ( $F$ )
2.   select  $i$  features with the biggest MI score
3.   get predictors matrix  $X^{**} \in R^{n \times i}$ 
4.   divide  $X^{**}$  and  $y^*$  on train/test subsets.
5.   create model object and fit with train data.
6.   calculate Acc, AUC on test subset
7.   IF (Acc  $\geq$  Acc_min) && (AUC  $\geq$  AUC_min)
8.     add model object to vector  $\vec{S}$ 
9.   ENDIF
10. ENDFOR

```

Fig. 1. Modeling algorithm

In lines 2–5, candidate models are fitted with increasing by one number of included features. The first candidate model includes one feature, and the last candidate includes F features, where F is the maximum number of features in our datasets. In line 2, i -features are sorted according to their mutual information (MI) score in descending order; the i -features are selected from the start of the sorted list with MI scores. In lines 3–4, the predictors' matrix is truncated to include only selected features, and a target variable and train and test subsets are formed from it. In lines 5–6, a model fits with the training subset, and performance metrics accuracy (Acc) and area under the ROC curve (AUC) are calculated on the test subset. In lines 7–8, if the model object's calculated performance satisfies the minimum performed level of accuracy (Acc_min) and AUC (AUC_min), then the model object is saved in the result vector \vec{S} .

A general guideline is used in supervised machine learning with the following intervals for accuracy and AUC metrics:

- if Accuracy/AUC = 0.5, then this is a guessing equal to flipping a coin;
- if $0.5 < \text{Accuracy/AUC} < 0.7$, then this is poor classification;
- if $0.7 < \text{Accuracy/AUC} < 0.8$, then this is an acceptable classification;
- if $0.8 < \text{Accuracy/AUC} < 0.9$, then this is an excellent classification;
- if Accuracy/AUC ≥ 0.9 , then this is outstanding discrimination.

The above rules are to be used to set minimum values of Accuracy and AUC for the algorithm (fig. 1). If in the result of the execution of the algorithm vector \vec{S} is empty, then we propose to lower the minimum values of the

performance metrics. If vector \vec{S} is not empty, then we can move on to grade candidate models by Bayes factor and grades the steps undertaken are described as per pseudocode (fig. 2).

For each model object from \vec{S} in line 2 is identified a BIC weight, denoted as w_m . Then in lines 3–6, each model is graded according to the Bays factor's rules. "Positive" models are saved to vector \vec{M}_1 . "Strong" – to vector \vec{M}_2 , "Very Strong" – to vector \vec{M}_3 . In current work, we ignored "weak" candidates.

```

Input:  $\vec{S}$ 
Output:  $\vec{M}_1, \vec{M}_2, \vec{M}_3$ 
1. FOR  $m$  in  $\vec{S}$ 
2.   calculate BIC  $w_m$ 
3.   IF  $w_m \geq 0.5$  &&  $w_m < 0.75$ 
4.     add the model object to  $\vec{M}_1$ 
5.   ELSEIF  $w_m \geq 0.75$  &&  $w_m < 0.95$ 
6.     add the model object to  $\vec{M}_2$ 
7.   ELSEIF  $w_m \geq 0.95$ 
8.     add the model object to  $\vec{M}_3$ 
9.   ENDIF
10. ENDFOR

```

Fig. 2. Steps to grade the candidate models

Model validation. The assessment of the goodness of fit of models from $\vec{M}_1, \vec{M}_2, \vec{M}_3$ compared to best-fit models \vec{B} was done through the steps as per pseudocode (fig. 3).

```

Input:  $\vec{M}_1, \vec{M}_2, \vec{M}_3, \vec{B}$ 
Output:  $\vec{R}, M$ 
 $\vec{R} = []$ 
1. calculate  $\bar{A}_2$  for models from  $\vec{B}$ .
2. calculate  $\bar{A}_1$  for models from  $\vec{M}_3$ .
3. calculate  $t$ -statistic.
4. compute the two-tailed  $p$ -value of the normal distribution.
5. IF  $p$ -value  $> 0.05$ 
6.   add the model object to  $\vec{R}$ 
7. ENDIF
8. IF  $\vec{R}$  is empty
9.   repeat steps 2–7 with models from  $\vec{M}_2$  and  $\vec{M}_1$ 
10. ENDIF
11. IF  $\vec{R}$  is empty
12.   select a model  $M$  with the best performance from  $\vec{M}_1, \vec{M}_2, \vec{M}_3$  or model from  $\vec{B}$ 
13. ENDIF

```

Fig. 3. Steps to assess goodness of fit of parsimonious models

In lines [1–4], mean values, the t -statistic, and the two-tailed p -value of the normal distribution for "very strong" models are computed. In lines 5–7, if the null hypothesis is not rejected, then a parsimonious model is added to the result vector \vec{R} . Lines 8–9 are executed if the goodness of fit test is failed for models from \vec{M}_3 . In this case, steps 2–7 are repeated with "strong" and "positive" models. Lines 11–14 are executed only if all models from $\vec{M}_1, \vec{M}_2, \vec{M}_3$ failed goodness of fit test. In that scenario, the model M with the best performance is selected from $\vec{M}_1, \vec{M}_2, \vec{M}_3$. The algorithm (fig. 3.) leaves experts to finally judge which model to use if all parsimonious candidate models failed the assessment. It could be either best-fit models from \vec{B} or the parsimonious model with the best performance metrics because their minimum values are set as an input parameter of the algorithm (fig. 2).

Study results and their discussion. Multiple candidate models are created according to designed algorithm (fig. 1). Applied Bayes factor grades as specified

in fig. 2 allowed to select: a "very strong" parsimonious model to recommend Interviews as an elicitation technique that included eight features and evaluated with performance Accuracy = 90 %; AUC = 98 % (fig. 4, *a*) which are 4 % and 1 % lower than Accuracy and AUC of best-fit model (table 2 – "Interviews"); a "very strong" parsimonious model to recommend Document analysis as an elicitation technique that included five features and evaluated with performance Accuracy = 90 %; AUC = 95 % (fig. 4, *b*) which are 1 % and 2 % lower than Accuracy and AUC of best-fit model (table 2 – "Document Analysis").

A "strong" parsimonious model to recommend Interface analysis as an elicitation technique that included nine features and evaluated with performance Accuracy = 81 %; AUC = 88 % (fig. 5, *a*) which are 3 % and 2 % lower than Accuracy and AUC of best-fit model (table 2 – "Interface Analysis"); a "strong" parsimonious model to recommend Process analysis as an elicitation technique that included fifteen features and evaluated with performance Accuracy = 81 %; AUC = 86 % (fig. 5, *b*) which are 1 %

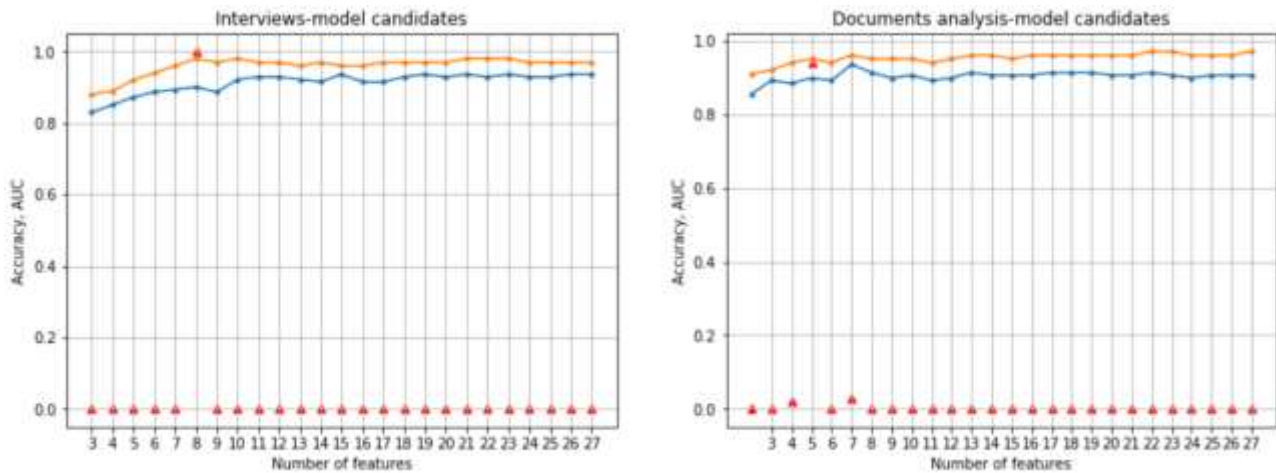


Fig. 4. Candidate model(s) BIC weight, Accuracy, AUC:
a – Interviews; *b* – Document analysis

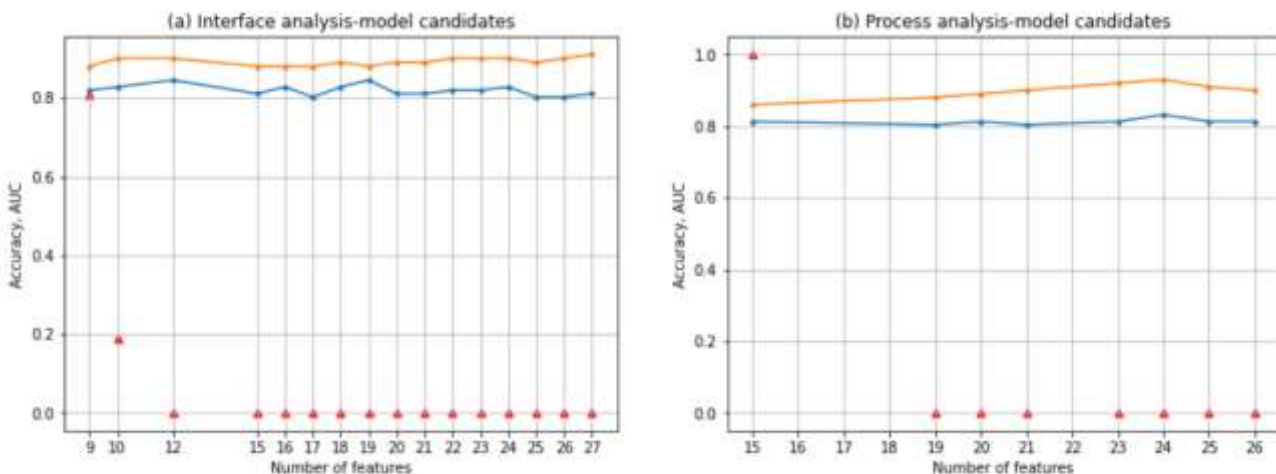


Fig. 5. Candidate model(s) BIC weight, Accuracy, AUC:
a – Interface analysis; *b* – Process analysis

Table 2 – Performance of parsimonious and best-fit models

Techniques	Accuracy			AUC		
	Parsimonious	Best-fit	Difference	Parsimonious	Best-fit	Difference
Interviews	0.90	0.94	–4 %	0.96	0.97	–1 %
Document Analysis	0.90	0.91	–1 %	0.95	0.97	–2 %
Interface Analysis	0.82	0.84	–3 %	0.88	0.90	–2 %
Process Analysis	0.81	0.82	–1 %	0.86	0.88	–2 %

and 2 % lower than Accuracy and AUC of best-fit model (table 2 – "Process Analysis").

As specified in fig. 3, the hypothesis test is applied with the models' performance metrics from table 2. A null hypothesis H_0 : the mean difference between the parsimonious and best-fit models' accuracies is 0. An H_1 hypothesis: the difference between the accuracies is different. T -statistic per equation 7 gives $t = -2.8$. The p -value with the degree of freedom equal to 3 is 0.066, which is greater than 0.05, so our H_0 is accepted, i.e., the parsimonious models are accepted, and best-fit can be ignored. Similarly, the hypothesis test with a null hypothesis H_0 : the mean difference between the values of AUC of the parsimonious model and the values of AUC of best-fit models is 0. H_1 hypothesis: the mean difference between AUC values is different. T -statistic per equation 7 gives $t = -7$. The p -value of $t = -7$ with the degree of freedom equal to 3 is 0.006, which is less than 0.05, so our H_0 is rejected, and the best-fit model is preferable due to the reduced parsimonious model's performance based on the mean value of AUC.

Thus, it can be concluded that applying the algorithm as per fig. 3 with each performance metric in sequence helps to identify when a parsimonious model's performance is degraded and decide on the suitable model's selection. We accepted the built parsimonious models in the current test experiment because the model's accuracy didn't deteriorate based on the goodness of fit test.

Conclusions and perspectives of further development. In the current study, the algorithms to build parsimonious candidate machine learning models and select the best candidate were designed and tested with four datasets collected for requirement elicitation technique selection. The results showed that the best candidate models graded as "very strong" and "strong" reduced the number of features: three times for Interviews and Interface analysis, five times for Document analysis, and 1.7 times for Process analysis. It helped to avoid the overfitting data problem.

The designed algorithm to assess the goodness of fit of the parsimonious models was applied with two performance metrics: accuracy and AUC in sequence. Based on the received results is concluded that by applying the proposed procedure, the gaps in the performance of the parsimonious model compared to the best-fit model can be detected, and a decision on the suitable model's selection can be made.

In summary, the obtained results allow us to recommend using a parsimonious model instead of the best-fit model to predict the using the particular elicitation

technique in IT projects and form recommendations based on that model.

Several directions for future research can be considered, such as creating machine learning models for other business analysis techniques, e.g., specification and modeling, prioritization, and structure of business analysis architecture.

References

- Gobov D., Huchenko, I. Influence of the Software Development Project Context on the Requirements Elicitation Techniques Selection. *Lecture Notes on Data Engineering and Communications Technologies*. Cham: Springer, 2021. Vol 83. P. 208–218. DOI: 10.1007/978-3-030-80472-5_18.
- Gobov D., Solovei O. Approaches to Improving the Accuracy of Machine Learning Models in Requirements Elicitation Techniques Selection. *arXiv:2303.14762*, 2023. DOI: 10.48550/arXiv.2303.14762.
- Harrell F. E. *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis*. New York, USA: Springer, 2001. 582 p.
- Anderson D., Burnham K. *Model selection and multi-model inference*. Second ed. New York, USA: Springer-Verlag, 2004. 488 p. DOI: 10.1007/b97636.
- Bursac Z., Gauss H. C., Williams D. K., and Hosmer D. W. Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine*. 2008. Vol. 3 (17). P. 3–17. DOI: 10.1186/1751-0473-3-17.
- Zhang Z. Variable selection with stepwise and best subset approaches. *Annals of translational medicine*. 2016. Vol. 4(7). P. 136. DOI: <https://doi.org/10.21037/atm.2016.03.35>.
- Solovei O. New organization process of feature selection by filter with correlation-based features selection method. *Innovative Technologies and Scientific Solutions for Industries*. 2022. Vol. 3 (21). P. 39–50. DOI: 10.30837/ITSSI.2022.21.039.
- Vrieze S. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological methods*. 2012. Vol. 17, № 2. P. 228–243. DOI: 10.1037/a0027127.
- Arnold T. Uninformative parameters and model selection using Akaike's Information Criterion. *The Journal of Wildlife Management*. 2010. Vol. 74, № 6. P. 1175–1178. DOI: 10.1111/j.1937-2817.2010.tb01236.x.
- van de Schoot R., Depaoli S., King R., Kramer B., Märtens K., Tadesse M. G., Vannucci M., Gelman A., Veen D., Willemsen J., Yau C. Bayesian statistics and modelling. *Nature Reviews Methods Primers*. 2021. Vol. 1. DOI: 10.1038/s43586-020-00001-2.
- Lesaffre E., Lawson A. *Bayesian Biostatistics*. West Sussex, United Kingdom: John Wiley & Sons, 2012. 544 p. DOI: 10.1002/9781119942412.
- Wasserman L. Bayesian model selection and model averaging. *Journal of mathematical psychology*. 2000. Vol. 44. № 1. P. 92–107. DOI: 10.1006/jmps.1999.1278.
- Rouder J.N., Speckman P.L., Sun D., Morey R. D., Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*. 2009. Vol.16. P. 225–237. DOI: 10.3758/PBR.16.2.225.
- Hosmer Jr. D., Lemeshov S., Sturdivant R. *Applied logistic regression*, West Sussex, United Kingdom: John Wiley & Sons, 2013. 510 p.

15. Ghoroghi A., Rezgui Y., Petri I., Beach T. Advances in application of machine learning to life cycle assessment: a literature review. *The International Journal of Life Cycle Assessment*. 2022. Vol. 27. P. 433–456. DOI: 10.1007/s11367-022-02030-3.

References (transliterated)

- Gobov D., Huchenko, I. Influence of the Software Development Project Context on the Requirements Elicitation Techniques Selection. *Lecture Notes on Data Engineering and Communications Technologies*. Springer, Cham, 2021, vol 83, pp. 208–218. DOI: 10.1007/978-3-030-80472-5_18.
- Gobov D., Solovei O. Approaches to Improving the Accuracy of Machine Learning Models in Requirements Elicitation Techniques Selection, *arXiv:2303.14762*, 2023. DOI: 10.48550/arXiv.2303.14762.
- Harrell F. E. *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression and Survival Analysis*. New York, USA, Springer, 2001. 582 p.
- Anderson D., Burnham K. Model selection and multi-model inference. Second ed. New York, USA, Springer-Verlag, 2004. 488 p. DOI: 10.1007/b97636.
- Bursac Z., Gauss H. C., Williams D. K., and Hosmer D. W. Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine*. 2008, vol. 3 (17), pp. 3–17. DOI: 10.1186/1751-0473-3-17.
- Zhang Z. Variable selection with stepwise and best subset approaches. *Annals of translational medicine*. 2016, vol. 4 (7), p. 136. DOI: 10.21037/atm.2016.03.35.
- Solovei O. New organization process of feature selection by filter with correlation-based features selection method. *Innovative Technologies and Scientific Solutions for Industries*. 2022, vol. 3 (21), pp. 39–50. DOI: 10.30837/ITSSI.2022.21.039.
- Vrieze S. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological methods*. 2012, Vol. 17, no. 2, pp. 228–243. DOI:10.1037/a0027127.
- Arnold T. Uninformative parameters and model selection using Akaike's Information Criterion. *The Journal of Wildlife Management*. 2010, vol. 74, no. 6, pp. 1175–1178. DOI: 10.1111/j.1937-2817.2010.tb01236.x.
- van de Schoot R., Depaoli S., King R., Kramer B., Märtens K., Tadesse M. G., Vannucci M., Gelman A., Veen D., Willemsen J., Yau C. Bayesian statistics and modelling. *Nature Reviews Methods Primers*. 2021, 1, vol. 1. DOI: 10.1038/s43586-020-00001-2.
- Lesaffre E., Lawson A. *Bayesian Biostatistics*. West Sussex, United Kingdom, John Wiley & Sons, 2012. 544 p. DOI: 10.1002/9781119942412.
- Wasserman L. Bayesian model selection and model averaging. *Journal of mathematical psychology*. 2000, vol. 44, no. 1, pp. 92–107. DOI: 10.1006/jmps.1999.1278.
- Rouder J.N., Speckman P.L., Sun D., Morey R. D., Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*. 2009, vol.16, pp. 225–237. DOI: 10.3758/PBR.16.2.225.
- Hosmer Jr. D., Lemeshov S., Sturdivant R. *Applied logistic regression*, West Sussex, United Kingdom, John Wiley & Sons, 2013, 510 p.
- Ghoroghi A., Rezgui Y., Petri I., Beach T. Advances in application of machine learning to life cycle assessment: a literature review. *The International Journal of Life Cycle Assessment*. 2022, vol. 27, pp. 433–456. DOI: 10.1007/s11367-022-02030-3.

Received 15.04.2023

УДК 004.415.25

О. Л. СОЛОВЕЙ, кандидат технічних наук (PhD), Київський Національний університет будівництва і архітектури, доцент кафедри інформаційних технологій проектування та прикладної математики, м. Київ, Україна; e-mail: solovey.ol@knuba.edu.ua; ORCID: <https://orcid.org/0000-0001-8774-7243>

Д. А. ГОБОВ, кандидат технічних наук, Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", доцент кафедри інформатики та програмної інженерії факультету інформатики та обчислювальної техніки, Київ, Україна; e-mail: d.gobov@kpi.ua; ORCID: <https://orcid.org/0000-0001-9964-0339>

ЕКОНОМНІ МОДЕЛІ МАШИННОГО НАВЧАННЯ ДЛЯ ВИБОРУ ТЕХНІК ВІЯВЛЕННЯ ВИМОГ

Предметом дослідження в статті є алгоритми машинного навчання, що використовуються для вибору технік виявлення вимог. Метою роботи є побудова ефективних економічних моделей машинного навчання для прогнозування використання методів виявлення вимог в IT-проектах, які дозволяють використовувати якомога менше незалежних змінних без значного погіршення якості прогнозу. У статті вирішуються наступні завдання: розробка алгоритму побудови економічних моделей-кандидатів машинного навчання для вибору техніки виявлення вимог на основі зібраної інформації про досвід практикуючих фахівців, оцінка точності моделі економічного машинного навчання та розробка алгоритму вибору найкращої моделі-кандидата. Використовуються такі методи: теорія алгоритмів, теорія статистики, методи вибірки, теорія моделювання даних та наукові експерименти. Було отримано наступні результати: 1) для вибору технік виявлення вимог побудовано економічні моделі-кандидати машинного навчання. Вони включали менше параметрів, що допомагає у майбутньому уникнути проблем із перенавчанням, пов'язаних із найкращими моделями; 2) відповідно до запропонованого алгоритму для відбору найкращого кандидата була обрана одна економічна модель із задовільною продуктивністю. Висновок. Запропоновано алгоритм для побудови ошадливих моделей-кандидатів для вибору техніки виявлення вимог, які дозволяють уникнути проблеми перенавчання. Алгоритм вибору найкращої моделі-кандидата визначає, коли продуктивність економічної моделі погіршується, і приймає рішення щодо вибору відповідної моделі. Обидва запропоновані алгоритми були успішно протестовані з чотирма наборами даних і можуть бути запропоновані для їх розширення для інших.

Ключові слова: техніки виявлення вимог, байєсівський інформаційний критерій (BIC), фактор Байєса, довірчий інтервал, економічна модель, точність, площа під кривою ROC.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Соловей Ольга Леонідівна, Solovei Olga Leonidivna

Автор 2 / Author 2: Гобов Денис Андрійович, Gobov Denys Andriyovych