http://doi.org/10.35784/iapgos.8053

received: 06.06.2025 | revised: 31.03.2026 | accepted: 23.04.2026 | available online: 30.06.2026

ANALYSIS OF THE INCREASE IN MODEL FORECASTING ACCURACY AFTER DATA NORMALIZATION

Vladyslav Pylypenko, Vladyslava Skidan, Antonina Volivach

Kyiv National University of Technologies and Design, Department of Information and Computer Technologies, Kyiv, Ukraine

Abstract. *In modern machine learning, data preprocessing is a crucial step that significantly affects the performance of classification models and the accuracy of prediction. This research investigates the impact of data normalization on the prediction accuracy of two common machine learning algorithms: logistic regression and support vector machine (SVM). Data normalization methods, including logarithmic transformation and scaling methods (such as RobustScaler and MinMaxScaler), were applied to evaluate their impact on model accuracy, balanced accuracy, F1 score, and area under the receiver operating characteristic (AUC). The results showed that the normalization process led to a significant improvement in model performance. In particular, logistic regression showed a moderate increase in all key metrics: accuracy improved by 14.6%, balanced accuracy by 12.3%, F1 score by 11.5%, and AUC by 3.9%. In contrast, SVM showed significant improvements: accuracy increased by 42.5%, balanced accuracy by 81.3%, F1 score by 21.3%, and AUC by 117.9%. These results highlight the importance of preprocessing steps for models sensitive to feature scaling, such as SVM. The ROC curves confirmed the improvement in classification performance, with the AUC of SVM increasing from 0.407 to 0.887 after normalization, indicating a shift from a low-performing model to one that more accurately discriminates between positive and negative classes. This improvement demonstrates the effectiveness of normalization in improving the generalization and robustness of the model, especially when handling unbalanced datasets. The result highlights that data normalization is an important step before training and learning a model. It not only improves the accuracy of the model, but also provides better adaptability and stability in real-world applications. And it is also an important step towards achieving reliable and accurate predictions. The results of the study highlight the importance of data preprocessing in machine learning processes, especially for algorithms that depend on feature scale or geometric methods, such as SVM and logistic regression.*

Keywords: AI models, ML algorithms, machine learning, data normalization, C#, Python

ANALIZA WZROSTU DOKŁADNOŚCI PROGNOZOWANIA MODELU PO NORMALIZACJI DANYCH

Streszczenie. *W nowoczesnym uczeniu maszynowym wstępne przetwarzanie danych jest kluczowym krokiem, który znacząco wpływa na wydajność modeli klasyfikacyjnych i dokładność prognozowania. W niniejszym badaniu zbadano wpływ normalizacji danych na dokładność prognozowania dwóch powszechnych algorytmów uczenia maszynowego: regresji logistycznej i maszyny wektorów nośnych (SVM). Metody normalizacji danych, w tym metody transformacji logarytmicznej i skalowania (takie jak RobustScaler i MinMaxScaler), zastosowano w celu oceny ich wpływu na dokładność modelu, zrównoważoną dokładność, wynik F1 i obszar pod charakterystyką operacyjną odbiornika (AUC). Wyniki pokazały, że proces normalizacji doprowadził do znacznej poprawy wydajności modelu. W szczególności regresja logistyczna wykazała umiarkowany wzrost wszystkich kluczowych metryk: dokładność poprawiła się o 14,6%, zrównoważona dokładność o 12,3%, wynik F1 o 11,5%, a AUC o 3,9%. Natomiast SVM wykazało znaczące ulepszenia: dokładność wzrosła o 42,5%, zrównoważona dokładność o 81,3%, wynik F1 o 21,3%, a AUC o 117,9%. Wyniki te podkreślają znaczenie kroków wstępnego przetwarzania dla modeli wrażliwych na skalowanie cech, takich jak SVM. Krzywe ROC potwierdziły poprawę wydajności klasyfikacji, przy czym AUC SVM wzrosło z 0,407 do 0,887 po normalizacji, co wskazuje na przejście od modelu o niskiej wydajności do takiego, który dokładniej rozróżnia klasy dodatnie i ujemne. Ta poprawa pokazuje skuteczność normalizacji w poprawie generalizacji i solidności modelu, szczególnie podczas obsługi niezrównoważonych zestawów danych. Wynik podkreśla, że normalizacja danych jest ważnym krokiem przed trenowaniem i nauką modelu. Nie tylko poprawia dokładność modelu, ale także zapewnia lepszą adaptowalność i stabilność w rzeczywistych zastosowaniach. Jest to również ważny krok w kierunku osiągnięcia wiarygodnych i dokładnych prognoz. Wyniki badania podkreślają znaczenie wstępnego przetwarzania danych w procesach uczenia maszynowego, zwłaszcza w przypadku algorytmów zależnych od skali cech lub metod geometrycznych, takich jak SVM i regresja logistyczna.*

Slowa kluczowe: modele sztucznej inteligencji, algorytmy uczenia maszynowego, uczenie maszynowe, normalizacja danych, C#, Python

Introduction

Data normalization is a preprocessing approach where data is either scaled or transformed to make each feature contribute equally. The output accuracy of machine learning algorithms depends on the quality of the data to obtain a generalized predictive model for a classification problem. In fact, data normalization is an important step before training machine learning models because it significantly affects the quality, stability, and speed of model learning [5]. Normalization refers to the reduction of numerical features to a common scale, for example, between 0 and 1 or with a zero mean and one standard deviation [2]. This is especially important when the data contains features with different units of measurement or values that differ by several orders of magnitude. Without normalization, the model may perceive features with a larger range of values as more important, even if they do not actually have a significant impact on the result. The lack of normalization is most detrimental to models that are based on distance or weighting, such as logistic regression, support vector methods, neural networks, and gradient descent algorithms [9]. In such cases, unnormalized features can lead to incorrect learning or very slow algorithm convergence. For example, logistic regression with regularization assumes that all features have approximately the same scale, otherwise the model will unfairly penalize those with larger values. In support vector methods, the scale of the features directly affects the formation of the separation hyperplane, and without normalization, the result may be inaccurate.

However, it should be noted that not all models are equally sensitive to the scale of the data. Decision trees, random forests, and naive Bayes are not dependent on the scale of the features, so normalization is not critical for them. When working with models that depend on distance metrics or weighting coefficients, data normalization is a mandatory practice that often significantly improves prediction accuracy. Therefore, investigating the impact of data normalization on the prediction results of machine learning models is a rather important task.

1. Literature review

Analysis of literature sources showed that the use of data normalization is one of the effective tools for improving the accuracy of forecasting models. And the use of modern information technologies, machine learning methods and an object-oriented approach have significantly simplified its implementation for research and use.

In the study [1], forecasting was performed using a logistic regression model. To study the effectiveness of this classifier, a dataset consisting of 1599 instances was applied to the model. Each record consists of 11 input attributes, which are determined by the dependence of the logistic regression model. Three types of normalization methods applied to the dataset were used:

Min-Max, Z-score and decimal scaling. As a result, the accuracy of correctly predicted values was 56.99%, 55.74%, 56.16% and 56.16% for the base-level model, Min-Max, Z-score and decimal scaling, respectively. Typically, it is expected that the prediction accuracy should improve after a normalized dataset, but in this case, we did not achieve a significant increase in accuracy. This behaviour may be related to the magnitude of the independent variables in the dataset, which are already close to each other before normalization.

The study presented in [13] investigates the impact of data scaling methods on the performance of machine learning models.. Linear models (SVM and LR) as well as ensemble models (RF, XGBoost and AdaBoost) were evaluated under three conditions: without data transformation, with standardized data and with normalized data. It was observed that linear models such as SVM and LR are more sensitive to scaling methods compared to ensemble models such as RF, XGBoost and AdaBoost. Linear models consistently struggled without scaling, especially for medium and small datasets. For small datasets, the accuracy of SVM increased from 61.23% without scaling to over 91% with scaling. LR showed similar trends, improving to 91.63% with scaling. The positive impact of scaling methods is most evident on medium-sized datasets with linear models. In some cases, both scaling methods showed overall satisfactory performance on all evaluation metrics. Therefore, it is important to make a choice based on context and characteristics.

A comparative study of the prediction performance of normalization methods using data mining classification was conducted in [10]. The aim of the study was to compare three normalization methods in terms of classification accuracy provided by normalized data: Z-score, decimal scaling, and statistical column. Six well-known classifications were used to evaluate the normalization methods: K-nearest neighbour, decision tree, artificial neural network, support vector method, naive Bayesian method, and binary logistic regression. Decimal scaling and decision tree classification had the highest accuracy of 79.2%. Using the statistical column in the K-nearest neighbour classification yielded an accuracy of 81.7%. Using the Z-score and statistical column, the classification using the naive Bayesian method yielded accuracies of 79.5% and 99.7%, respectively. However, when using decimal scaling, the K-nearest neighbour classification had the maximum accuracy of 100%. Of all the normalization and classification methods, the statistical column and K-nearest neighbour classification showed the highest accuracy performance.

2. Methodology

Two machine learning models were used in the study: one based on Logistic Regression and the Support Vector Machine (SVM) method [6]. The choice of these algorithms is due to their efficiency in classification tasks, the ability to work with a large number of features, as well as their prevalence in practical applications. The models were built in the PyCharm development environment in the Python programming language. The scikit-learn library was used to build the models [3]. The following libraries were used to build the graphs: seaborn and matplotlib [15]. Before training the models, data preprocessing was performed, including normalization. The data for training and training the models were data on students' academic performance, attendance, and their interaction with video materials with a total of more than 2.5 thousand records [7]. A fragment of the table with input data is presented in Table 1.

To parse the data into training and educational datasets, an application was written in the C# programming language in the Microsoft Visual Studio development environment. The CsvHelper library was used to process the table columns. Since most models are sensitive to the scale of features, especially those that use distances or weights to calculate solutions, without proper normalization, the model can focus on features with a larger range of values, ignoring others, which actually leads to distortion of the results and loss of accuracy.

For logistic regression, two main normalization methods were used: logarithmic transformation and scaling via RobustScaler. The logarithmic transformation (safe_log1p function) allows to reduce the influence of large values, to make the distribution more symmetrical and linearly suitable for analysis. Mathematically, the log transformation can be described as follows [4]:

$$x' = \log(x + 1) \quad (1)$$

where x is the original feature value to be transformed. It can be any number, including zero; x′ is the new, transformed value.

$$x' = (x - \text{median}(X))/(\text{IRQ}(X)) \quad (2)$$

where x is the original feature value to be transformed; X is a vector or column of data containing all values of a particular feature; median(X) is the median of all values in X, i.e. the central value that divides the sample in half; IRQ(X) is the Interquartile Range, calculated as IQR(X) = Q3 - Q1.

Two scaling approaches were used for the SVM-based model: MinMaxScaler and RobustScaler. The first scales all features to the range [0, 1], which is a standard approach for SVM models, especially when using an RBF (radial basis function) as the kernel. Since distances in vector space directly affect the similarity calculation, a mismatch in feature scales can completely destroy the learning process. Mathematically, MinMax scaling can be described as [11]:

$$x' = (x - x\_min)/(x\_max - x\_min) \quad (3)$$

where x is the original value of the feature that needs to be normalized; x_min is the minimum value of this feature in the entire sample; x_max is the maximum value of the same feature in the sample; x΄ is the normalized value scaled to the interval [0,1].

As in the case of logistic regression, RobustScaler was additionally used to protect against outliers, especially with a large number of observations or unpredictable variance. Special attention was paid to the issue of the impact of normalization on the accuracy of the models. For this purpose, a comparison of the prediction results on unnormalized and normalized data was carried out. It was found that without normalization, the accuracy of SVM did not exceed 41%, that is, it was actually equivalent to random prediction in binary classification. After applying transformations (logarithmic and scaling), the accuracy increased.

The quality assessment of the constructed classification models was carried out on the basis of a set of metrics that allow comprehensively characterizing their effectiveness in the context of a specific task. The main criteria were accuracy, balanced accuracy, sensitivity, specificity, and F1-measure [12]. Accuracy reflects the total proportion of correctly classified examples among all predictions and is one of the basic indicators of effectiveness. However, it may not be sufficiently informative

Table 1. Fragment of a table with input data

id	DiscMark	LectVisit	PractVisit	LabVisit	TotalVisit	Duration	PlayCount	PauseCount	StopCount	Complete
6	82	83	-1	83	83	27	1	1	0	1
7	51	25	-1	33	29	0	0	0	0	0

Note: id is a unique user identifier; DiscMark – grade for discipline; LectVisit – percentage of lecture attendance; PractVisit – percentage of attendance at practical classes; LabVisit – the percentage of attendance at laboratory classes; TotalVisit – total attendance percentage; Duration – video viewing duration (minutes); PlayCount – the number of clicks on the Play button; PauseCount – the number of Pause button clicks; StopCount – the number of clicks on the Stop button; Complete – the status of completion of watching the video to the end (1 – yes, 0 – no).

in cases where the data has a significant imbalance between the number of objects of different classes. In such cases, it is advisable to use balanced accuracy, which takes into account both the correctness of the classification of the positive class (sensitivity) and the negative (specificity), allowing to obtain a more objective assessment of the model. Sensitivity reflects the ability of the model to correctly identify all objects of the positive class, that is, it shows how well the model responds to positive cases [14]. Specificity, in turn, demonstrates the ability of the model to correctly identify negative cases and not confuse them with positive ones. Both indicators are critically important in cases where the cost of misclassifying one of the classes is significantly higher. The F1-measure allows for a more balanced assessment of the quality of the model, especially in conditions of uneven distribution of classes or asymmetric importance of false positive and false negative results. ROC-curves were also constructed for each model, illustrating the relationship between sensitivity and frequency of false positive results at different decision thresholds. The area under the ROC-curve (AUC-ROC) serves as a summary characteristic of the model's ability to distinguish classes: the closer this area is to 1, the better the overall classification quality. Special attention was paid to the influence of data preprocessing – in particular, normalization of input features – on the quality of the models. As a result of comparing the two approaches: with and without normalization, it was found that the model with pre-normalized features demonstrates better performance in most metrics, in particular, an increase in the F1-measure, balanced accuracy and a decrease in the number of false positive predictions. The shape of the ROC curve also improved and the AUC value increased, indicating a more stable operation of the model with variations in the classification threshold [8].

Thus, the analysis confirmed the importance of data preprocessing in the process of building machine learning models, especially when using algorithms that are sensitive to the scale of features, such as logistic regression. Normalization not only improves the speed and stability of model training, but also increases its ability to generalize to new data, which is a key factor for successful application in real conditions.

3. Results

For a detailed assessment of the quality of the classification model built on the basis of logistic regression, a comparative analysis of two variants of the model was conducted – without prior normalization of the input data and with its application. For this purpose, the corresponding confusion matrices were formed, which reflect the number of correct and incorrect predictions for each class. They allow us to clearly assess the impact of preprocessing (normalization) on the behaviour of the model, in particular on the change in the distribution of errors between classes. The resulting confusion matrices are presented in Fig. 1 and Fig. 2.

The calculation of such key characteristics of the models (based on logistic regression), as accuracy, balanced accuracy, sensitivity, specificity and F1 Score, is presented in Table 2. These indicators allow us to assess the efficiency of the classification model and its ability to correctly identify both positive and negative cases in the sample.

To visually assess the quality of classification models, in particular the ability to distinguish classes at different decision thresholds, ROC curves were constructed for prediction models, which are presented in Fig. 3-4. The curve allows us to analyse how the ratio between sensitivity (True Positive Rate) and the frequency of false positives (False Positive Rate) changes when the threshold at which the model classifies observations as positive or negative changes.

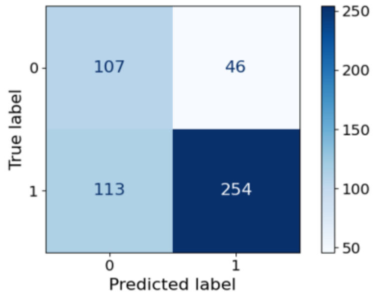

Fig. 1. Confusion matrices for logistic regression models before data normalization

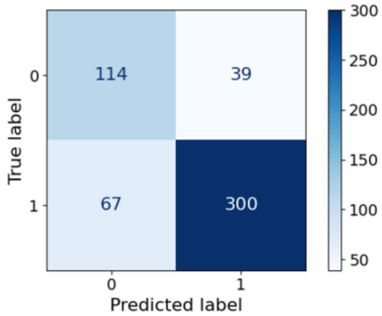

Fig. 2. Confusion matrices for logistic regression models after data normalization

Table 2. Calculation of model performance indicators based on logistic regression without/with data normalization

Model performance	Without normalization	With normalization
Accuracy	0.694	0.796
Balanced Accuracy	0.696	0.781
Sensitivity	0.692	0.817
Specificity	0.847	0.885
F1 Score	0.762	0.850

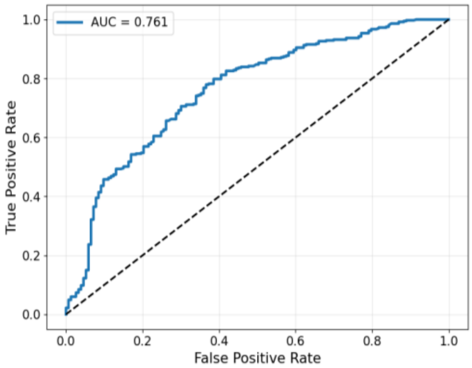

Fig. 3. ROC curve of the logistic regression model before data normalization

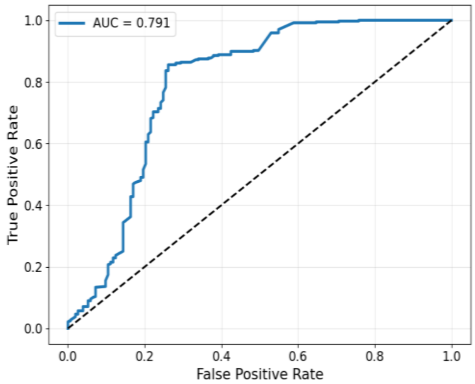

Fig. 4. ROC curve of the logistic regression model after data normalization

The plots show that before data normalization, the Area Under the Curve was 0.761. This indicates that the model had a moderate ability to distinguish between positive and negative classes, but a certain number of false classifications were still present. The curve passed above the diagonal line of random classification, which already indicates the presence of useful information in the features, but did not provide high accuracy at different thresholds. After performing data normalization, in particular, logarithmic transformation and scaling using RobustScaler, which reduced the impact of outliers and brought the features to the same scale, the AUC value increased to 0.79. This means that the model began to better distinguish between objects of the two classes at different threshold values. The ROC curve became higher and closer to the upper left corner of the graph, which reflects an increase in sensitivity without a significant increase in false positives. It can be concluded that data normalization had a positive impact on classification quality, making the model more robust to feature-scale imbalances and reducing the impact of unevenly distributed values, but the overall gain is not significant. For a detailed assessment of the quality of the classification model based on the support vector method, error matrices were constructed for two models (without and with data normalization), which are presented in Fig. 5 and Fig. 6.

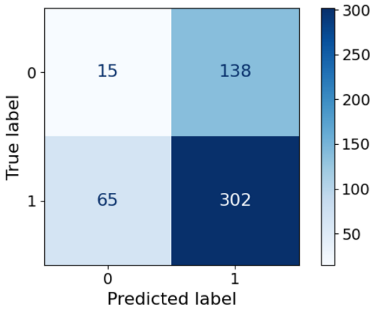

Fig. 5. Confusion matrices for models based on the support vector method before data normalization

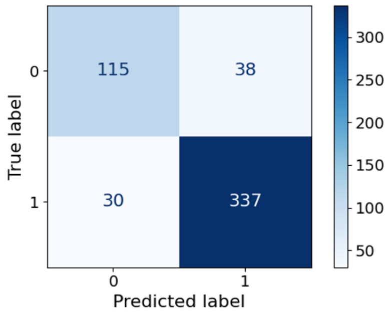

Fig. 6. Confusion matrices for models based on the support vector method after data normalization

The calculation of such key characteristics of models (based on the support vector method), as accuracy, balanced accuracy, sensitivity, specificity and F1 Score, is presented in Table 3.

Table 3. Calculation of model performance indicators based on the support vector method without/with data normalization

Model performance	Without normalization	With normalization
Accuracy	0.610	0.869
Balanced Accuracy	0.460	0.835
Sensitivity	0.823	0.918
Specificity	0.686	0.899
F1 Score	0.748	0.908

To visually assess the quality of classification models, in particular the ability to distinguish classes at different decision thresholds, ROC curves were constructed for the prediction models, which are presented in Fig. 7-8.

In the ROC curve plot constructed for the model before data normalization, the area under the curve was 0.407. This is below the threshold for random classification (0.5), which indicates that the model actually classified incorrectly more often than by chance. The ROC curve in this case passed below the diagonal line, which visually indicates that the model makes the opposite assumptions – often mistakenly recognizing negative examples as positive, and vice versa. After applying normalization, which included scaling using the MinMaxScaler scaler, the model began to perform better, and the AUC increased significantly to 0.887. This already indicates a high quality of classification, as the curve rose up and became much closer to the upper left corner of the graph, indicating greater sensitivity of the model while maintaining a low level of false positive decisions. The curve looks smoother with a clear separation from the random diagonal, which visually confirms a significant improvement. After data normalization, the model received balanced input parameters, which allowed it to build a correct classification boundary. This emphasizes the critical importance of data preprocessing for obtaining reliable and effective results in machine learning. The overall accuracy gain for key values characterizing the efficiency and quality of the model is presented in Table 4.

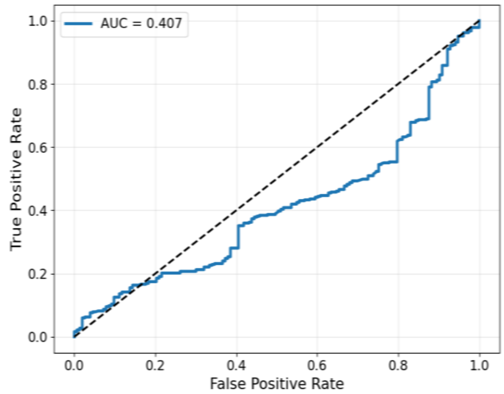

Fig. 7. ROC curve of the logistic regression model before data normalization

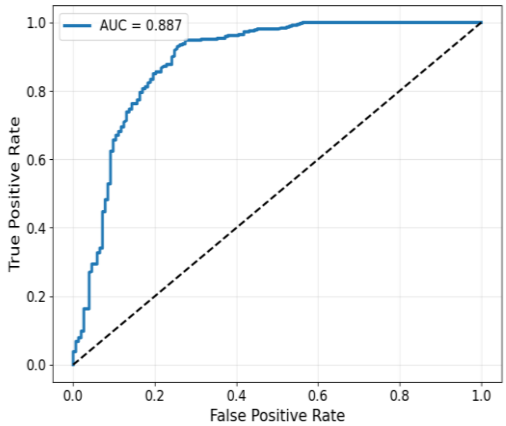

Fig. 8. ROC curve of the logistic regression model before data normalization

Table 4. Comparison of the main indicators of model accuracy gain after data normalization

Model performance	Logistic Regression	Support Vector Method
Accuracy	+14.68%	+42.59%
Balanced Accuracy	+12.30%	+81.33%
F1 Score	+11.59%	+21.36%
Area Under Curve	+3.96%	+117.91%

The logistic regression model showed a moderate but consistent improvement in all key metrics after normalization. Accuracy increased by 14.6%, indicating an overall improvement in the model's predictions. Balanced accuracy, which takes into account both classes in the case of imbalance, increased by 12.3%, indicating an improvement in the model's ability to correctly classify both positive and negative examples. An 11.5% increase in F1 score indicates an improvement in the balance between accuracy and sensitivity. The AUC (area under the receiver operating characteristic curve) improved by 3.9%, confirming a small but real improvement in the model's ability to separate classes at different thresholds. The support vector machine (SVM) model showed a significant improvement in almost all metrics after normalization. Accuracy increased by 42.5%, which is a significant indicator of improved prediction quality. The largest increase is observed in balanced accuracy +81.3%, which is especially important in problems with uneven class distribution. This means that after normalization, the SVM began to effectively process both classes, without predominating one of them. The F1 Score improved by 21.3%, which indicates an increase in the harmonious balance between accuracy and completeness. The largest is the increase in AUC by 117.9%, which demonstrates a radical improvement in the model's ability to separate classes – from almost uninformative to high-quality classification. Thus, it can be argued that data normalization significantly improved the performance of both models, but the SVM-based model was much more sensitive to preprocessing. Properly chosen normalization is a critical condition for achieving high accuracy and stability of machine learning models.

4. Conclusions

1. Data normalization is a critical preprocessing step in classification tasks, especially for models sensitive to the scale of input features, such as the support vector machine (SVM). The use of approaches such as logarithmic transformation and scaling (RobustScaler, MinMaxScaler) stabilizes the feature distribution, reduces the impact of outliers, and improves the generalization quality of models.
2. Analysis of classification quality indicators before and after normalization demonstrated a statistically significant increase in the efficiency of models. In particular, for logistic regression, the improvement was: accuracy: +14.6%, balanced accuracy: +12.3%, F1 score: +11.5%, AUC: +3.9%. For SVM, the increase was: +42.5%, +81.3%, +21.3%, and +117.9%, respectively. This indicates an increase in both the resolution of the models and their ability to correctly classify under different data types.
3. Analysis of the constructed ROC curves confirmed the increase in the ability of the models to separate positive and negative classes. In particular, the AUC for SVM increased from 0.40 to 0.88 after normalization, which indicates a radical improvement in the qualitative characteristics of the model and the correction of the initial overfitting or incorrectly trained classification.
4. The improvement in the F1-measure value after normalization demonstrates an increase in the balance between accuracy and completeness, which is especially important when working with unbalanced classes. This avoids situations where the model overestimates one of the classes and provides a more uniform classification quality.
5. The results of the study confirm the theoretical justification for the need for preliminary transformation of features when using algorithms based on distance or geometric separation (for example, SVM or logistic regression). Without such a stage, models may not only be inaccurate, but also incorrect from the point of view of decision-making.

References

[1] Adeyemo, N. A., Wimmer, H., & Powell, L. M. (2018). Effects of Normalization Techniques on Logistic Regression in Data Science. *2018 Proceedings of the Conference on Information Systems Applied Research*, *11*, 1–9.

[2] Ali, P. J. M., & Faraj, R. H. (2014). Data Normalization and Standardization: A Technical Report. *Machine Learning Technical Reports*, *1*(1), 1–6. https://doi.org/10.13140/RG.2.2.28948.04489

[3] Bisong, E. (2019). Introduction to Scikit-learn. In E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform* (pp. 215–229). Apress. https://doi.org/10.1007/978-1-4842-4470-8_18

[4] Booeshaghi, A. S., & Pachter, L. (2021). Normalization of single-cell RNA-seq counts by log(x + 1) or log(1 + x). *Bioinformatics*, *37*(15), 2223–2224. https://doi.org/10.1093/bioinformatics/btab085

[5] Jo, J.-M. (2019). Effectiveness of Normalization Pre-Processing of Big Data to the Machine Learning Performance. *The Journal of the Korea Institute of Electronic Communication Sciences*, *14*(3), 547–552. https://doi.org/10.13067/JKIECS.2019.14.3.547

[6] Musa, A. B. (2013). Comparative study on classification performance between support vector machine and logistic regression. *International Journal of Machine Learning and Cybernetics*, *4*(1), 13–24. https://doi.org/10.1007/s13042-012-0068-x

[7] Pylypenko, V., Statsenko, V., Bila, T., & Statsenko, D. (2024). Determining the influence of data on working with video materials on the accuracy of student success prediction models. *Eastern-European Journal of Enterprise Technologies*, *5*(4 (131)), 52–62. https://doi.org/10.15587/1729-4061.2024.313333

[8] Raju, V. N. G., Lakshmi, K. P., Jain, V. M., Kalidindi, A., & Padma, V. (2020). Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 729–735. https://doi.org/10.1109/ICSSIT48917.2020.9214160

[9] Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, *97*, 105524. https://doi.org/10.1016/j.asoc.2019.105524

[10] Sinsomboonthong, S. (2021). Efficiency Comparison in Prediction of Normalization with Data Mining Classification. *Advances in Science, Technology and Engineering Systems Journal*, *6*(4), 130–137. https://doi.org/10.25046/aj060415

[11] Sinsomboonthong, S. (2022). Performance Comparison of New Adjusted Min-Max with Decimal Scaling and Statistical Column Normalization Methods for Artificial Neural Network Classification. *International Journal of Mathematics and Mathematical Sciences*, *2022*, 1–9. https://doi.org/10.1155/2022/3584406

[12] Statsenko, V., & Pylypenko, V. (2024). Assessment of the efficiency of the success prediction model using machine learning methods. *Herald of Khmelnytskyi National University. Technical Sciences*, *331*(1), 271–276. https://doi.org/10.31891/2307-5732-2024-331-41

[13] Sujon, K. M., Hassan, R. B., Towshi, Z. T., Othman, M. A., Samad, M. A., & Choi, K. (2024). When to Use Standardization and Normalization: Empirical Evidence From Machine Learning Models and XAI. *IEEE Access*, *12*, 135300–135314. https://doi.org/10.1109/ACCESS.2024.3462434

[14] Ting, K. M. (2017). Sensitivity and Specificity. In C. Sammut & G. I. Webb (Eds), *Encyclopedia of Machine Learning and Data Mining* (pp. 1152–1152). Springer US. https://doi.org/10.1007/978-1-4899-7687-1_758

[15] Waskom, M. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

M.Sc. Vladyslav Pylypenko
e-mail: software.proger@gmail.com

Ph.D. student of the Department of Computer Science, Kyiv National University of Technologies and Design, Ukraine. Developer of desktop and mobile applications, specializing in the implementation of machine learning. Senior lecturer at the Department of Information and Computer Technologies.
Scientific interests: computer engineering, processing and analysis of big data, work with artificial intelligence models, software development.

https://orcid.org/0000-0002-2761-4817

Ph.D Vladyslava Skidan
e-mail: skidan.vv@knutd.edu.ua

Candidate of Technical Sciences, associate professor, Head of the Department of Information and Computer Technologies, Kyiv National University of Technologies and Design, Ukraine. Author of over 100 scientific papers.
Scientific interests: computer engineering and technology, big data processing and analysis, software development.

https://orcid.org/0000-0002-8358-9759

Ph.D. Antonina Volivach
e-mail: volivach.ap@knutd.edu.ua

Candidate of Technical Sciences, associate professor of the Department of Information and Computer Technologies, Kyiv National University of Technologies and Design, Ukraine. Author of over 100 scientific papers.
Scientific interests: computer engineering and technology, big data processing and analysis, software development.

https://orcid.org/0000-0002-7119-7774