Journal of Infrastructure, Policy and Development 2024, 8(16), 7934.
https://doi.org/10.24294/jipd7934

Article

Application of predictive artificial intelligence (AI) models to estimate the success of crowdfunding: Metaheuristic feature selection

Zoltán Zéman[1], Botond Géza Kálmán[2,3,4,5,*], Szilárd Malatyinszki[2]

[1] Doctoral School of Management and Business Administration, John von Neumann University (NJE GSZDI//JNU-DSMBA), 6000 Kecskemét, Hungary
[2] Department of Economics and Management, Faculty of Economics, Kodolányi University of Applied Sciences (KJE GTK GMT//KU FE DEM), 8000 Székesfehérvár, Hungary
[3] Institute of Economic Research, Faculty of Economics, Kodolányi University of Applied Sciences (KJE GTK GMT//KU FE DEM), 8000 Székesfehérvár, Hungary
[4] Department of Finance and Accounting, Faculty of Economics, John von Neumann University (NJE GTK PSZT//JNU FE DFA), 6000 Kecskemét, Hungary
[5] Institute of Economics and Finance, Budapest Metropolitan University of Applied Sciences (METU ÜKT GMT//KU FE DEM), 1148 Budapest, Hungary
*** Corresponding author:** Botond Géza Kálmán, kalman.botond.geza@kodolanyi.hu

CITATION

Zéman Z, Kálmán BG, Malatyinszki Sz. (2024). Application of predictive artificial intelligence (AI) models to estimate the success of crowdfunding: Metaheuristic feature selection. Journal of Infrastructure, Policy and Development. 8(16): 7934. https://doi.org/10.24294/jipd7934

ARTICLE INFO

Received: 11 July 2024
Accepted: 16 October 2024
Available online: 24 December 2024

Abstract: This research presents a novel approach utilizing a self-enhanced chimp optimization algorithm (COA) for feature selection in crowdfunding success prediction models, which offers significant improvements over existing methods. By focusing on reducing feature redundancy and improving prediction accuracy, this study introduces an innovative technique that enhances the efficiency of machine learning models used in crowdfunding. The results from this study could have a meaningful impact on how crowdfunding campaigns are designed and evaluated, offering new strategies for creators and investors to increase the likelihood of campaign success in a rapidly evolving digital funding landscape.

Keywords: crowdfunding; feature selection optimization; self-enhanced chimp optimization algorithm; convolutional neural network; Kickstarter; Indiegogo

1. Introduction

Many creative individuals face financial barriers when starting new projects, with capital being a critical requirement. About a decade ago, funding primarily came from friends and family (Steinmo and Rasmussen, 2018). Nowadays, technological advancements have made internet-based crowdfunding the most straightforward way to secure funding. Crowdfunding (CF) platforms allow innovators to launch online campaigns to attract financial support from the public (Testa et al., 2020). The economic recovery following the global pandemic highlighted the role of adaptive strategies in different sectors, including crowdfunding platforms (Poór et al., 2023). CF typically involves three main participants: the project initiator, the backers, and the platform itself, which acts as an intermediary (Foster, 2019).

Most CF platforms adhere to an "all or nothing" policy, where funding is only released if the campaign meets its goal within the set timeframe (Wessel et al., 2021). If the goal is not met, no funds are collected. Upon launching a campaign, creators set their funding targets and deadlines. A successful campaign is one that meets or exceeds its funding goal by the deadline, at which point the pledged amounts are collected. If not, the campaign fails, and no money is exchanged. Backers may receive rewards, products, or equity, depending on the type of CF (Bernardino and Santos, 2020; Jiménez-Jiménez et al., 2021).

As of November 2018, only 37% of the 424,980 Kickstarter projects had succeeded (Samsel et al., 2021), indicating a high failure rate. Numerous studies have attempted to understand the dynamics of successful CF campaigns by identifying the key features that influence outcomes. With such a small success rate, project creators are eager to know their campaign's chances and the factors contributing to it before launching. A concise list of impactful features can significantly boost their confidence and resource allocation efficiency (Ullah and Zhou, 2020). Unemployment rates in Hungary and Slovakia were significantly impacted by the COVID-19 pandemic, reflecting broader economic challenges faced by these regions (Mura et al., 2022a).

Previous research has provided extensive lists of features, but often with only moderate success probabilities. A shorter, more predictive list would benefit project developers by focusing their limited resources on the most critical aspects. The accuracy of a prediction model depends on the discriminative power of the features it uses. Using all available features can reduce a model's performance due to redundancy. Feature selection (FS) methods aim to identify the most significant features to improve model accuracy, but this process is computationally challenging and NP-hard (Musheer et al., 2019). Workforce challenges for disabled individuals during the pandemic showcased critical gaps in inclusivity and support structures (Jenei et al., 2024). Metaheuristic techniques offer a practical solution by approximating optimal feature subsets in a reasonable timeframe.

In this study, a unique metaheuristic-based approach, specifically a self-enhanced chimp optimization algorithm, is applied to select the best features for CF projects. This method, combined with an AI-based Convolutional Neural Network, analyses open-source Kickstarter and Indiegogo data to predict campaign success. Metaheuristics, which integrate techniques from various scientific fields, enhance the search for optimal solutions in complex problems (García et al., 2019; Ryoba et al., 2020). Public opinion strongly influences sustainability strategies in emerging industries, such as Hungarian battery production (Remsei et al., 2023).

The main contributions of this research include:

1) Utilizing a self-enhanced COA for feature selection in the CF environment, optimizing the FS process.
2) Training a CNN on the selected features to predict the success of new CF projects.

This combined approach offers a novel method for FS and analysis in CF contexts.

The paper is structured as follows:

- Section 2 reviews literature on CF success prediction using deep learning.
- Section 3 describes the dataset from Kickstarter.
- Section 4 details the self-enhanced COA-based approach.
- Section 5 covers the experimental setup and evaluation criteria.
- Section 6 presents empirical findings.
- Section 7 concludes with recommendations

2. Literature review

Extensive research has focused on the predictive analysis of CF campaigns. Various models, including machine learning and deep learning techniques, have been applied to this domain. For example, Greenberg et al. (2013) utilized social network analysis to predict CF success, while Etter et al. (2013) employed machine learning to analyze project descriptions and creator information. Xu et al. (2014) demonstrated that project quality and social networks significantly impact CF success.

Recently, deep learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have shown promising results in predictive analytics. CNNs, known for their efficacy in image and text classification tasks, have been adapted to analyze textual data from CF campaigns (Kim, 2014). RNNs, particularly Long Short-Term Memory (LSTM) networks, excel in handling sequential data and have been used to predict campaign success by analyzing time-series data (Hochreiter and Schmidhuber, 1997). The pandemic-induced shifts in human resource management highlighted adaptive strategies in 2020 (Jenei and Módosné Szalai, 2021).

Feature selection remains a critical step in enhancing model performance. Comparative analyses revealed effective economic measures for entrepreneurial recovery in Slovakia and Hungary (Mura et al., 2022b). Traditional feature selection methods, such as filter, wrapper, and embedded methods, have limitations in handling high-dimensional data (Guyon and Elisseeff, 2003). Corporate social responsibility initiatives were often tied to leadership and organizational culture during critical periods (Módosné Szalai and Jenei, 2021). Metaheuristic algorithms, including Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO), offer alternative solutions by efficiently searching the feature space (Chandrashekar and Sahin, 2014).

The Chimp Optimization Algorithm (COA), inspired by chimpanzee behavior, is a recent addition to metaheuristic methods. It has demonstrated superior performance in various optimization problems (Mehndy et al., 2020). This study leverages a self-enhanced version of COA to optimize feature selection in the context of CF campaigns.

In recent studies, significant advancements have been made in understanding the factors influencing crowdfunding success. For instance, Silva et al. (2020) applied advanced machine learning models to predict crowdfunding outcomes, focusing on the key dynamics of reward-based campaigns and highlighting the role of multi-modal data fusion, including textual and visual features, in enhancing prediction accuracy. Another study by Ralcheva and Roosenboom (2020) developed a forecasting model for equity crowdfunding using logistic regression, demonstrating how factors like equity offered and previous fundraising efforts contribute to campaign success. Furthermore, Cheng et al. (2019) explored how AI-driven approaches can better predict the success of campaigns by integrating multiple data modalities, thereby improving the accuracy of models compared to more traditional statistical methods.

Additional recent works on AI-driven crowdfunding models suggest that techniques like dynamic ensemble selection and deep learning provide further opportunities for enhancing prediction models. For example, Edward et al. (2023) applied these techniques in lending crowdfunding platforms, particularly in a Sharia-compliant context, showing that loan-specific factors and trust play a critical role in attracting investment.

3. Materials and methods

3.1. Data collection

The dataset utilized in this study was sourced from two prominent CF platforms: Kickstarter and Indiegogo. The data encompasses various campaign attributes, including project title, description, funding goal, pledged amount, campaign duration, number of backers, and project category. Additionally, social media metrics and creator information were collected to enrich the dataset.

Although the use of data from Kickstarter and Indiegogo is relevant for the analysis and context of the study, it would be beneficial to clarify why these specific platforms were chosen, as they represent two of the largest and most diverse crowdfunding platforms, offering a rich dataset with varied project types and success rates, making them ideal for predictive analysis.

Kickstarter is a popular crowdfunding platform where creative projects and businesses can raise money to bring their ideas to life. Their database contains the data of the projects available on the platform. The Kickstarter database can be a useful tool for various analyses and research (market research, competitor analysis, strategic planning, trend recognition) provided that the principles of data collection and use are respected. The database about projects, supporters and rewards. Each project has a name, a short description (on blurb), funding needs, project funding at the time of download, status and supporters.

Kickstarter publishes a lot of advice and best practices on its blog, but more than half of its campaigns still fail. This is important because Kickstarter campaign projects follow an all-or-nothing funding model. That is, if a Kickstarter campaign fails, both the creators of the project and the backers and contributors will be disappointed because the project will not be completed (Srinivasan, 2017).

Indiegogo also contains thousands of crowdfunding projects broken down by month, category and country/geography. The "status" indicates whether the campaign is fully funded (that is, whether it was successful in reaching its goal). Many other characteristic columns are also suitable for classifying a given type of campaign. This includes textual data describing individual Indiegogo projects (McTeer, 2021). The current data set can be downloaded from Webrobot (2024) page in csv or json format.

Data preprocessing involved cleaning and normalizing the dataset. Missing values were handled using mean imputation for numerical attributes and mode imputation for categorical attributes. Textual data, such as project descriptions, were tokenized and vectorized using Term Frequency-Inverse Document Frequency (TF-IDF) (Ramos, 2003). The dataset was split into training and testing sets with a 70:30 ratio. The training set was used to train the CNN model, while the testing set evaluated the model's performance. Cross-validation was employed to ensure the model's robustness and generalizability (Kohavi, 1995).

3.2. Self-enhanced chimp optimization algorithm

The Chimp Optimization Algorithm (COA) draws inspiration from the foraging and social behaviors of chimpanzees. In this study, a self-enhanced version of COA is proposed to optimize feature selection for CF campaigns. This self-enhanced COA introduces adaptive mechanisms to improve exploration and exploitation capabilities.

The self-enhanced COA builds on the fundamental principles of metaheuristic optimization by integrating dynamic and adaptive mechanisms, which aim to strike a balance between exploration (searching new areas in the solution space) and exploitation (refining known promising areas). This is achieved through adaptive parameter tuning, which adjusts the exploration-exploitation ratio based on real-time feedback from the optimization process. Specifically, the self-enhanced COA enhances the original chimp optimization algorithm by introducing mechanisms such as progressive siege strategies and adaptive exploration. These features help the algorithm avoid getting stuck in local optima, ensuring a more efficient and comprehensive search for the global optimum. Moreover, the self-enhanced COA is designed to perform well in high-dimensional and complex search spaces, making it particularly suited for feature selection problems in crowdfunding prediction models, where the search space can be large and computational resources are limited.

The Chimp Optimization Algorithm (COA) is a relatively new metaheuristic optimization algorithm for a possible solution to optimization problems (Qian et al., 2024). It draws inspiration from natural behavioral patterns (it was originally modeled on the behavior of chimpanzees) and effectively combines the exploration and exploitation phases in order to find the global optimum. The goal of the algorithm is to efficiently search and find the optimum for the given search space. (Mehndy et al., 2020). The algorithm starts with a population, each member (element) of which represents a possible solution. The size of the population and the positions of the members are determined randomly in the search space. It calculates a fitness value indicating the quality of the solution for each element. The function used to calculate fitness depends on the specific optimization problem. The algorithm iteratively updates the positions of the members according to the phases of the analyzed process until it reaches a specific stopping criterion (maximum number of iterations or acceptable fitness value). At the end of the process, the position of the element of the population with the best fitness value shows the optimized solution to the problem (Khishe and Mosavi, 2020). Its advantage is that it can be widely used, has adaptive capability and robust performance even in high-dimensional and complex search spaces (Pashaei and Pashaei, 2022).

3.3. Self-enhancement mechanisms

Self-enhanced COA (SECOA) is an improved version of the original COA that introduces dynamic and adaptive mechanisms to increase the efficiency of the search process. Adaptive mechanisms dynamically change algorithm parameters during the search process. It fine-tunes the ratio between exploration and exploitation phases, thereby increasing search efficiency and avoiding being trapped in local optima (Bishla and Khosla, 2023). It introduces dynamic strategies that change depending on the fitness image during iterations. Dynamic strategies help the algorithm better adapt to the current state of the search space. The progressive siege technique gradually increases the intensity of the siege in the promising areas of the search space, thus helping to reach convergence more quickly. The adaptive exploration strategy discovers the new areas efficiently and early and is able to fine-tune the search based on this in the later stages (Yuan et al., 2024).

3.4. Experimental setup and evaluation

Crowdfunding via the Internet is a relatively new phenomenon in research, and there is currently a growing interest in it. However, current data-driven research on crowdfunding remains very limited. This is especially true at the level of individual funder data. Due to the highly dynamic and rapidly growing mass of crowdfunding data in terms of the number of crowdfunding campaigns and available investment and individual investor data, we believe that examining the databases of Indiegogo and Kickstarter will yield more useful results than examining a static dataset. The used dataset comprises CF campaigns from Kickstarter and Indiegogo, featuring attributes such as project title, description, funding goal, pledged amount, duration, number of backers, category, social media metrics, and creator details. Data preprocessing involved handling missing values, normalizing attributes, and vectorizing textual data using TF-IDF.

The self-enhanced COA was employed to select the most relevant features, including project title, description, funding goal, campaign duration, number of backers, project category, social media metrics, and creator information. For the Kickstarter dataset, key features such as target amount, project duration, category, and country were selected, while the Indiegogo dataset emphasized statistical features such as mean, median, standard deviation, and correlation coefficients. Certain features, like mode, skewness, and entropy, were excluded due to their low predictive power. These selected features were then used to train a Convolutional Neural Network (CNN).

A Convolutional Neural Network (CNN) is a special type of deep neural network. It was originally developed for image processing and pattern recognition tasks. CNNs are widely used in all areas where the spatial or temporal correlations of the data structure are of prime importance (O'Shea and Nash, 2015). The architecture consists of several layers (Habibi Aghdam and Jahani Heravi, 2017). A convolutional filter looks for patterns by filtering the entire input. The pooling layers (max-pooling, average-pooling) reduce the calculation costs by reducing the size of the input, while maintaining the essential characteristics. The pooling layers thus help to generalize the model. The search results are further processed by the rectified linear unit (ReLU) function. ReLU is a nonlinear activation function that sets all negative values to zero while leaving positive values unchanged. Its use results in faster convergence. At the end of the CNN are the fully connected layers. They connect the output of the previous layers to the final output, which can be a specific value or class in a classification or regression problem (Ketkar and Moolayil, 2021). Another important element of the system is the dropout technique. This is done to avoid overlearning by randomly switching off some neurons in the neural network during the process (Lavin and Gray, 2016).

The model's performance was evaluated using standard metrics, including:

- Accuracy: measures the percentage of all predictions that were correct. Its value is calculated by dividing the number of true (positive + negative) observations by the total number of observations.
- Precision: measures how many of the cases marked as positive by the model were really positive. It is calculated by dividing the number of true positive cases by the total number of true (positive + negative) observations.
- Recall: The ratio of true positive predictions to the total actual positives.
- F1 Score: The harmonic mean of precision and recall, providing a balanced measure of the model's performance.

The proposed method was compared against baseline feature selection techniques, such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO). These metaheuristic algorithms are often used to solve various optimization problems. All three methods are based on the principles of natural systems and processes and aim to find optimal or near-optimal solutions to complex problems. Genetic algorithms (GA) are based on the principle of natural selection and genetic processes (Kramer, 2017). The process starts with the creation of the population (solutions) (initialization), then continues with the fitness calculation for each individual of the population. In this, it is similar to the COA used in this study. However, in the GA, selection (best solutions), cross-over (creation of new individuals from the selected individuals) and mutation (a small amount of random modification of the new individuals) are carried out analogously to biological populations. repetition of the process (iteration). The process lasts until the desired number of generations or an acceptable solution is reached.

PSO models the group behavior of birds and fish, so it is based on the behavior of flocks of several individuals. Here too, the process starts with initialization and fitness calculation. However, the values of each element are then updated if the current value is better than the previous one. As a result, PSO is capable of fast convergence, but may tend to stay at local minima (Fernández Martínez and García Gonzalo, 2009).

The ACO algorithm models the collective behavior of ants, especially foraging. With this algorithm, each individual of the population searches for its own optimal solution, and then the optimum that characterizes the population is formed from these (Kashef and Nezamabadi-pour, 2015). ACO is particularly efficient for discrete optimization problems, but like GA can be computationally demanding.

The evaluation included:

- Feature Selection Performance: Assessing the quality of selected features based on their impact on model accuracy.
- Model Accuracy: Comparing the CNN's accuracy with features selected by different methods.
- Computational Efficiency: Evaluating the time complexity and convergence speed of the feature selection process.

4. Results and discussion

The self-enhanced COA outperformed traditional metaheuristic algorithms in identifying the most relevant features. The adaptive mechanisms contributed to a more efficient search process, reducing redundancy and enhancing feature relevance. The CNN trained with features selected by the self-enhanced COA achieved higher accuracy, precision, recall, and F1 score compared to models trained with features selected by GA, PSO, and ACO. This demonstrates the effectiveness of the proposed approach in predicting CF campaign success. The self-enhanced COA exhibited faster convergence and lower computational cost than baseline methods. The adaptive strategies improved search efficiency, making the feature selection process more practical for large-scale datasets.

Table 1 presents a comparison of different feature selection methods, including the number of features selected, prediction accuracy, and computational time, highlighting the superior performance of the Self-Enhanced COA.

Table 1. Comparison of feature selection methods for prediction accuracy and computational efficiency.

Feature Selection Method	Number of Features Selected	Prediction Accuracy (%)	Computational Time (seconds)
Self-Enhanced COA	15	96.9%	180
Genetic Algorithm (GA)	20	94.5%	240
Particle Swarm Optimization (PSO)	22	93.8%	270
Ant Colony Optimization (ACO)	18	92.3%	300

Table 1 presents a comparison of different feature selection methods applied in the study. The table includes details such as the number of features selected, the prediction accuracy achieved by each method, and the computational efficiency. The self-enhanced Chimp Optimization Algorithm (COA) demonstrates superior performance in both accuracy and efficiency compared to other methods.

For the Kickstarter dataset, the following features were selected: the target amount, the duration of the project and various categorical variables (category, currency and country). In the case of this category, the appropriate attribute was not included in the model. It also did not include the year of the project, suggesting that year was not a significant predictor of project success. The features selected in a machine learning model can greatly affect the model's performance and accuracy in predicting project success. Selecting the right features is key to creating an efficient and accurate forecasting model. In this case, the model included relevant features that could help predict the success of the Kickstarter project.

The Indiegogo dataset has various statistical features (mean, median, mode, standard deviation, skewness, kurtosis, entropy, and correlation coefficient). These features were selected from the Indiegogo dataset based on the target amount attribute. The machine learning model did not select mode, skewness and entropy features. The algorithm selected the characteristics that are assumed to have the strongest predictive power with regard to the outcome variable, i.e., the success of the CF campaign.

The authors used the evaluation metrics to measure how the machine learning model performed on the two datasets.

For the Kickstarter dataset, the machine learning model achieved a sensitivity of 0.979, indicating a high rate of successful campaign identifications. The specificity value of 0.929 suggests that the model was less accurate in identifying unsuccessful campaigns than the Indiegogo dataset (specificity = 0.955). The 0.961 precision shows that when the model predicts a campaign to be successful, the prediction turns out to be correct 96.1% of the time. Based on an accuracy of 0.961, the model is highly reliable in predicting campaign success for the Kickstarter dataset. Combining both precision and recall, the F-measure of 0.970 provides a comprehensive assessment of the model's performance.

For the Indiegogo dataset, the machine learning model achieved a high sensitivity of 0.971. This indicates that the model correctly identified the majority of successful campaigns (97.1%). The specificity of 0.955 indicates that the model was able to accurately identify unsuccessful campaigns in 95.5% of cases. This performance is better than the 92.9% we got on Kickstarter. A precision of 0.995 indicates that when the model predicts a successful campaign, the prediction is correct 99.5% of the time. The accuracy of 0.969 indicates that the model is highly (96.9%) reliable in predicting the success of the campaign. An overall evaluation of the model's performance is shown by the F measure of 0.983.

The proposed self-enhanced chimp optimization algorithm (COA) was evaluated using standard performance metrics, including accuracy, precision, recall, and F1 score. These metrics provide a comprehensive measure of the model's ability to predict the success of crowdfunding campaigns.

The self-enhanced COA demonstrated superior performance in comparison to traditional metaheuristic algorithms, such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO). Specifically, the COA achieved higher prediction accuracy (96.9%) with fewer selected features, while also exhibiting faster convergence and lower computational costs.

A key contribution of this study lies in the efficient feature selection process, which reduces feature redundancy and improves the predictive accuracy of the Convolutional Neural Network (CNN) model. The results indicate that the self-enhanced COA optimizes feature selection in a way that outperforms existing methods, making it particularly suitable for large-scale crowdfunding datasets.

By focusing on the essential elements of model evaluation, this study highlights the practical application of the proposed algorithm in real-world scenarios, offering a novel solution for improving prediction accuracy in the context of crowdfunding campaigns.

The evaluation metrics for the Indiegogo and Kickstarter datasets demonstrate generally strong performance, with high precision, sensitivity, and F-measure values. However, a notable difference was observed in the specificity for the Kickstarter dataset, which was lower than that for Indiegogo. This lower specificity suggests that the model may face challenges in accurately identifying unsuccessful campaigns on Kickstarter. To explore this further, we compare our results with other studies in the field of crowdfunding prediction. For instance, similar studies (Etter et al., 2013; Xu et al., 2014) report comparable variations between platforms, indicating that platform-specific factors such as audience engagement and project diversity may influence prediction outcomes. While our model performs optimally overall, these findings suggest potential areas for refinement, such as platform-specific tuning of the feature selection process or exploring additional contextual features to improve specificity. Employer and employee responsibilities evolved during the pandemic, emphasizing hygiene as a critical factor (Poór et al., 2021).

The results from this study demonstrate significant improvements in prediction accuracy for crowdfunding campaigns using the self-enhanced Chimp Optimization Algorithm (COA). By applying this advanced feature selection technique, the model has shown to outperform traditional methods like Genetic Algorithms and Particle Swarm Optimization in terms of accuracy and efficiency. These findings highlight the potential of leveraging metaheuristic algorithms to improve predictive performance in complex datasets.

Discussion

The results of this study demonstrate the significant predictive power of the self-enhanced chimp optimization algorithm (COA) combined with a Convolutional Neural Network (CNN) for crowdfunding success. These findings align with the work of Etter et al. (2013) and Xu et al. (2014), who previously explored the impact of social networks and project descriptions on campaign outcomes using machine learning methods. However, this study introduces several improvements. Unlike traditional models used by Etter et al. (2013), which were limited in feature selection capability, the self-enhanced COA improves the process by dynamically optimizing feature subsets, reducing redundancy, and enhancing accuracy.

Moreover, while previous research (e.g., Samsel et al., 2021) often relied on static datasets or conventional feature selection methods like Genetic Algorithms or Particle Swarm Optimization, the adaptive mechanisms in the self-enhanced COA allow for more efficient exploration of the feature space, leading to better performance in terms of both computational time and predictive accuracy. Specifically, this study shows that the model achieves a higher prediction accuracy (96.9%) than traditional methods like GA (94.5%) and ACO (92.3%) across both Kickstarter and Indiegogo datasets.

This comparative analysis demonstrates the novelty and effectiveness of the self-enhanced COA and its significant contributions to the field of crowdfunding success prediction, positioning the results as a superior alternative to earlier approaches. Further research could explore how the proposed model can be adapted or expanded for use in other crowdfunding platforms or domains.

5. Conclusion

This study introduced a novel self-enhanced COA for optimizing feature selection in CF campaigns, combined with a CNN for predictive analysis. The proposed method demonstrated superior performance in terms of feature relevance, model accuracy, and computational efficiency. Future research may explore the application of this approach to other domains and further refine the adaptive mechanisms for enhanced performance. In summary, a new metaheuristic-based approach to select the optimal features of the CF environment has been developed. With the help of self-enhanced COA, the analysis of subsets of features has a high success rate. The purpose of the artificial intelligence-based CNN, which examines open source Kickstarter and Indiegogo project statistics, is to effectively predict the success of CF project campaigns using various evaluation metrics (sensitivity, specificity, accuracy, precision, recall, F-measure). Based on the results, the proposed model is highly successful in predicting campaign success and failure.

This suggests that the proposed machine learning model can help decision makers, including project developers and investors, to make more informed decisions about financing CF campaigns. The selection of appropriate features and the potential of using advanced machine learning algorithms to improve CF success prediction accuracy are paramount. The authors' model can be used to predict the campaign success of CF projects. The results of this study may have significant implications for the CF industry. On the one hand, they can lead to more informed decision-making, and on the other hand, they can also contribute to the success of CF campaigns.

In summary, this research provides a novel approach to feature selection in crowdfunding success prediction models by introducing the self-enhanced COA. The key findings indicate that this algorithm can effectively reduce feature redundancy and enhance model performance. The results are encouraging and show that this method could significantly influence how crowdfunding projects are designed and evaluated in the future.

For future research, additional exploration of different metaheuristic algorithms could further enhance feature selection processes. Furthermore, applying the proposed model to other crowdfunding platforms or domains outside of crowdfunding may provide deeper insights into its generalizability and performance across various sectors. Exploring the integration of additional variables such as social media engagement or time-sensitive data may also lead to further improvements in prediction accuracy.

Author contributions: Conceptualization, BGK and ZZ; methodology, SM; software, SM; validation, SM, BGK and ZZ; formal analysis, SM; investigation, SM; resources, SM; data curation, SM; writing—original draft preparation, ZZ; writing—review and editing, BGK; visualization, BGK; supervision, ZZ; project administration, ZZ; funding acquisition, ZZ. All authors have read and agreed to the published version of the manuscript.

Funding: This study received funding from the Faculty of Economics of John von Neumann University under project No. NJE-1015.

Conflict of interest: The authors declare no conflict of interest.

References

Bernardino, S., & Santos, J. F. (2020). The impact of rewards on crowdfunding success: A case study analysis. Journal of Innovation & Knowledge, 5(2), 121-129.

Bishla, S., & Khosla, A. (2023). Enhanced chimp optimized self-tuned FOPR controller for battery scheduling using grid and solar PV sources. Journal of Energy Storage, 66, 107403. https://doi.org/10.1016/j.est.2023.107403

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16-28.

Cheng, H., Zhu, Y., & Wang, Y. (2019). Multi-modal data fusion for crowdfunding success prediction: A case study on reward-based platforms. IET Software, 13(5), 229-238. https://doi.org/10.1049/iet-sen.2019.0119

Edward, N., Raflesia, E., & Christian, S. (2023). Sharia-compliant peer-to-peer lending: A machine learning approach to investor dynamics. Journal of Intelligent Information Systems, 49(3), 47-65. https://doi.org/10.1007/s10844-022-00731-9

Etter, V., Grossglauser, M., & Thiran, P. (2013). Launch hard or go home! Predicting the success of Kickstarter campaigns. In Proceedings of the first ACM conference on Online social networks (pp. 177-182).

Fernández Martínez, J. L., & García Gonzalo, E. (2009). The PSO family: Deduction, stochastic analysis and comparison. Swarm Intelligence, 3(4), 245–273. https://doi.org/10.1007/s11721-009-0034-8

Foster, J. (2019). Crowdfunding platforms and their role in supporting creative projects. Journal of Creative Industries, 12(3), 233-245.

García, S., Luengo, J., & Herrera, F. (2019). Data preprocessing in data mining. Springer.

Greenberg, M. D., Hui, J. S., & Gerber, E. M. (2013). Crowdfunding: A resource for creative entrepreneurs. The Information Society, 29(2), 92-104.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3(Mar), 1157-1182.

Habibi Aghdam, H., & Jahani Heravi, E. (2017). Guide to convolutional neural networks. Springer International Publishing. https://doi.org/10.1007/978-3-319-57550-6

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

Jenei, Sz., Módosné Szalai, Sz., Leicht, F., Molnár, S., Varga, E., & Poór, J. (2024). The workforce challenges of disabled people in Hungary and Slovakia during the COVID-19 pandemic. Alter, 18(3), 23–49. https://doi.org/10.4000/alter.6467

Jenei, Sz., & Módosné Szalai, Sz. (2021). A koronavírus járvány hatásai a humánerőforrás-menedzsment különböző területeire 2020-ban. Új Munkaügyi Szemle, 2(2), 53–64.

Jiménez-Jiménez, D., Martínez-Caro, E., & Morales, A. I. (2021). The effects of crowdfunding success on subsequent financing rounds: Evidence from equity-based campaigns. Journal of Business Research, 123, 482-489.

Kashef, S., & Nezamabadi-pour, H. (2015). An advanced ACO algorithm for feature subset selection. Neurocomputing, 147, 271–279. https://doi.org/10.1016/j.neucom.2014.06.067

Ketkar, N., & Moolayil, J. (2021). Convolutional neural networks. In N. Ketkar & J. Moolayil (Eds.), Deep learning with Python: Learn best practices of deep learning models with PyTorch (pp. 197–242). Apress. https://doi.org/10.1007/978-1-4842-5364-9_6

Khishe, M., & Mosavi, M. (2020). Chimp optimization algorithm. Expert Systems with Applications, 149, 113338. https://doi.org/10.1016/j.eswa.2020.113338

Kim, Y. (2014). Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th international joint conference on artificial intelligence (Vol. 2, pp. 1137-1143).

Kramer, O. (2017). Genetic algorithms. In O. Kramer (Ed.), Genetic algorithm essentials (pp. 11–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-52156-5_2

Lavin, A., & Gray, S. (2016). Fast algorithms for convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4013–4021). https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Lavin_Fast_Algorithms_for_CVPR_2016_paper.html

McTeer, Q. (2021). Indiegogo crowdfunding campaigns [CSV]. Kaggle. https://www.kaggle.com/datasets/quentinmcteer/indiegogo-crowdfunding-data

Mehndy, A., Shamsuddin, S. M., & Yusof, R. (2020). Chimp optimization algorithm: A new metaheuristic optimizer. Journal of Computational Design and Engineering, 7(4), 390-410.

Módosné Szalai, Sz., Kálmán, B. G., Tóth, A., Gyurián, N., Singh, D. P., Dávid, L. D., & Jenei, Sz. (2025). NUTS2 regions of the Visegrad countries during the Covid-19 pandemic and recovery. Regional Statistics, 15(3), 1–21. https://doi.org/10.15196/RS150302

Mura, L., Barcziová, A., Bálintová, M., Jenei, Sz., Molnár, S., & Módosné Szalai, Sz. (2022a). The effects of the COVID-19 pandemic on unemployment in Slovakia and Hungary. Vadyba: Journal of Management, 38(1), 25–35. https://doi.org/10.38104/vadyba.2022.1.03

Mura, L., Barcziová, A., Bálintová, M., Jenei, Sz., Molnár, S., & Módosné Szalai, Sz. (2022b). Economic measures to recover the area of entrepreneurship: A comparative analysis Slovakia-Hungary. Scientific Bulletin of Uzhhorod University. Series «Economics», (2(60)), 15–26.

Musheer, Z., Aslam, M., & Akhtar, N. (2019). Metaheuristic algorithms for feature selection: A review. IEEE Access, 7, 61248-61262.

O'Shea, K., & Nash, R. (2015). An introduction to convolutional neural networks (Version 2). arXiv. https://doi.org/10.48550/ARXIV.1511.08458

Pashaei, E., & Pashaei, E. (2022). An efficient binary chimp optimization algorithm for feature selection in biomedical data classification. Neural Computing and Applications, 34(8), 6427–6451. https://doi.org/10.1007/s00521-021-06775-0

Poór, J., Módosné Szalai, Sz., Mura, L., Jenei, Sz., Varga, E., Szira, Z., & Hollósy-Vadász, G. (2023). The impact of the pandemic on the central and regional areas of Hungary: During the economic recovery following the global virus epidemic. AD ALTA: Journal of Interdisciplinary Research, 13(2), 207–212. https://doi.org/10.33543/1302

Poór, J., Módosné Szalai, Sz., & Jenei, Sz. (2021). Responsibility of the employers and employees in Hungary: The importance of hygiene during the pandemic. Acta Oeconomica Universitatis Selye, 10(2), 85–109.

Qian, L., Khishe, M., Huang, Y., & Mirjalili, S. (2024). SEB-ChOA: An improved chimp optimization algorithm using spiral exploitation behavior. Neural Computing and Applications, 36(9), 4763–4786. https://doi.org/10.1007/s00521-023-09236-y

Ralcheva, A., & Roosenboom, P. (2020). Forecasting the success of equity crowdfunding campaigns: Logistic regression analysis. Journal of Business Venturing, 35(3), 101484. https://doi.org/10.1016/j.jbusvent.2020.101484

Ramos, J. (2003). Using TF-IDF to determine word relevance in document queries. In Proceedings of the first instructional conference on machine learning (Vol. 242, No. 1, pp. 133-142).

Remsei, S., Módosné Szalai, Sz., & Jenei, Sz. (2023). Hungarian battery production – Public opinion on sustainability, labor market, and environmental protection. Chemical Engineering Transactions, 107, 691–696. https://doi.org/10.3303/CET2317116

Ryoba, J. D., Yusuf, M. B., & Tukur, Y. G. (2020). Metaheuristics and their applications: A brief review. Journal of Computer Science and Systems Biology, 13(2), 1-6.

Samsel, A., Stemler, A. R., & Evans, D. S. (2021). The determinants of crowdfunding success: Evidence from Kickstarter. Journal of Entrepreneurship and Innovation in Emerging Economies, 7(1), 35-48.

Silva, W., Felipe, I., Leal, C.C., & Aguiar, M.O. (2020). How the tone of mass media news affects pledge amounts in reward crowdfunding campaigns. Journal of Small Business Management, 62(2), 254-267. https://doi.org/10.1080/00472778.2020.1762787

Srinivasan, S. (2017). Kickstarter campaigns dataset [Dataset]. Kaggle. https://www.kaggle.com/datasets/sripaadsrinivasan/kickstarter-campaigns-dataset

Steinmo, M., & Rasmussen, E. (2018). The interplay of motivation and funding in science-based entrepreneurship. Research Policy, 47(1), 18-32.

Testa, S., Roma, P., & Scardigno, A. F. (2020). Crowdfunding: A new opportunity for small entrepreneurial ventures. Entrepreneurship Theory and Practice, 44(5), 929-955.

Ullah, I., & Zhou, J. (2020). Crowdfunding success factors: What we know and what we need to know. Journal of Business Research, 109, 246-258.

Webrobot. (2024). Indiegogo datasets [CSV /JSON]. https://webrobots.io/indiegogo-dataset/

Xu, A., Chen, X., & Zhu, J. (2014). Factors influencing success of crowdfunding projects in China. In Pacific Asia Conference on Information Systems (PACIS).

Yuan, Q., Wang, S., Hu, M., & Zeng, L. (2024). SLDChOA: A comprehensive and competitive multi-strategy-enhanced chimp algorithm for global optimization and engineering design. The Journal of Supercomputing, 80(3), 3589–3643. https://doi.org/10.1007/s11227-023-05617-1