sustainability

Brief Report

Benchmarking Biologically-Inspired Automatic Machine Learning for Economic Tasks

Teddy Lazebnik [1,*], Tzach Fleischer [2] and Amit Yaniv-Rosenfeld [3,4,5]

1 Department of Cancer Biology, Cancer Institute, University College London, London WC1E 6BT, UK
2 Department of Computer Science, Holon Institute of Technology, Holon 5810201, Israel
3 Shalvata Mental Health Care Center, Hod Hasharon 45100, Israel
4 Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv 6997801, Israel
5 Department of Management, Bar-Ilan University, Ramat-Gan 529002, Israel
* Correspondence: lazebnik.teddy@gmail.com

Abstract: Data-driven economic tasks have gained significant attention in economics, allowing researchers and policymakers to make better decisions and design efficient policies. Recently, with the advancement of machine learning (ML) and other artificial intelligence (AI) methods, researchers can now solve complex economic tasks with previously unseen performance and ease. However, to use such methods, one is required to have a non-trivial level of expertise in ML or AI, which currently is not standard knowledge in economics. In order to bridge this gap, automatic machine learning (AutoML) models have been developed, allowing non-experts to efficiently use advanced ML models with their data. Nonetheless, not all AutoML models are created equal in general, particularly for the unique properties associated with economic data. In this paper, we present a benchmarking study of biologically inspired and other AutoML techniques for economic tasks. We evaluate four different AutoML models alongside two baseline methods using a set of 50 diverse economic tasks. Our results show that biologically inspired AutoML models (slightly) outperformed non-biological AutoML in economic tasks, while all AutoML models outperformed the traditional methods. Based on our results, we conclude that biologically inspired AutoML has the potential to improve our economic understanding while shifting a large portion of the analysis burden from the economist to a computer.

Keywords: automatic machine learning; economy–biology interactions; data-driven economic tasks; evolutionary computation

check for updates

Citation: Lazebnik, T.; Fleischer, T.; Yaniv-Rosenfeld, A. Benchmarking Biologically-Inspired Automatic Machine Learning for Economic Tasks. *Sustainability* **2023**, *15*, 11232. https://doi.org/10.3390/su151411232

Academic Editors: Alan Randall, Peiwan Wang, Jiayin Qi and Feng Liu

Received: 19 May 2023
Revised: 11 July 2023
Accepted: 17 July 2023
Published: 19 July 2023

1. Introduction

Data-driven economic tasks have become increasingly popular in economics, allowing researchers and policymakers to make better decisions and design efficient policies [1–5]. Currently, most economic studies from a broad range of subjects are still using classical computational methods such as descriptive statistics [6–8], linear regression [9–11], rule-based models [12,13], and moving average prediction [14,15].

With the rapid development of machine learning (ML) and other artificial intelligence (AI) methods, researchers can now solve complex economic tasks with previously unseen performance levels [16]. However, utilizing these methods requires a significant level of expertise, which is currently time- and resource-consuming to obtain [17–19]. This challenge is, unfortunately, not unique to economics [20–23]. Nonetheless, it is particularly important in economics, as a wide range of economic tasks require the understanding and modeling of numerical patterns, a challenge in which ML models typically excel [24–26]. In particular, sustainability projects require careful and accurate analysis and planning, with a large number of parameters and processes to execute correctly [27,28]. To bridge this gap, automatic machine learning (AutoML) solutions have been developed, which allow non-experts to use advanced ML models on their data with ease [29–31].

Recent studies have shown promising usage of AutoML in economic settings [32]. For instance, the authors of [33] presented an AutoML framework to predict the casualty rate and direct economic loss induced by earthquakes that automates five processes: data collection, data preprocessing, ML algorithm and hyperparameter tuning, economic loss prediction, and model analysis. The authors showed that the model produced by their framework outperformed other ML models developed by experts on the same dataset. The authors of [34] investigated the performance of an AutoML methodology in forecasting bank failures, resulting in overwhelming performance with an area under the receiver operating characteristic curve (AUC) of 0.985. The authors of [35] introduced an integrative method that integrates data envelope analysis and AutoML to assess and predict performance in the Sustainable Development Goals (SDGs). The authors showed that this approach outperformed other analysis approaches for SDG data when the features did not correlate well with the target feature. The authors of [36] showed that for biological homeostasis systems, a biological regulatory system faces the objectives of being effective while using its resources in a responsible manner. The authors used Pareto optimal analysis from the economic realm, showing that the processes occurring in a cell population agree with resource allocation models from the economic domain. This outcome highlights the deep connection between the two disciplines. Recently, the authors of [37] showed that as part of the evolution of cells, a cell population utilizes a gambling strategy that is close to the optimal strategy predicted by a model designed for investment in highly uncertain markets.

Since it seems that biology and economics share a deep connection between them at the dynamic level, in this study, we aim to investigate if biologically inspired AutoML models would outperform other methods for data-driven economic tasks, focusing on static and time series regression tasks with at least several hundred samples. For this purpose, we benchmark three groups of models—traditional, biologically inspired AutoML, and other AutoML models—on a large and diverse set of data-driven economic regression tasks. We first reproduce, in scale, that AutoML is able to outperform traditional methods. In addition, we show that, on average, biologically inspired AutoML models outperform other AutoML models, supporting the line of work associating the biological and economic domains from a computational point of view. Thus, the novelty of the proposed work lies in the statistical analysis of a large number of economic tasks on a wide range of data-driven computation methods, first showing an empirical evaluation of AutoML usage in economics in general and that biologically inspired AutoML models in particular are usually favorable.

The rest of this paper is structured as follows. In Section 2, we review the literature connecting economic and biological modeling, followed by biologically inspired algorithms and automatic machine learning methods. In Section 3, we describe the datasets, models, and statistical analysis used in this research. Next, in Section 4, we present the results obtained from our experiments. In Section 5, we analyze and discuss the obtained results and offer concluding remarks with possible future work.

2. Related Work

Next, we briefly review the central computational methods pertaining to our challenge: biologically inspired algorithms and automatic machine learning.

2.1. Biologically Inspired Algorithms

Biologically inspired algorithms, which draw inspiration from biological and ecological processes to solve optimization problems, have gained popularity in recent years [38–40]. Several examples of such algorithms are the genetic [41], particle swarm optimization [42], and artificial immune system algorithms [43]. These algorithms presumably mimic the processes of natural selection, swarm behavior, and immune response to find optimal solutions to complex problems through an iterative process. Specifically, scholars observe organisms performing complex computational tasks and record how they perform them. Afterward, these records are broken down and modeled into several abstract computation steps that can be performed by a computer. Commonly, by performing these steps in the right order, the organisms are able to get closer to the desired goal(s), essentially defining iterative optimization algorithms [44].

Indeed, biologically inspired algorithms have gained much popularity in many fields, including engineering [45], physics [46], sociology [47], and others [48,49]. In the field of economics, biologically inspired algorithms have shown promising results in solving complex economic tasks, such as stock market prediction [50], portfolio optimization [51], and demand forecasting [52]. For instance, a genetic algorithm with a neural network model (i.e., also a biologically inspired model) has been used as a stock trading decision support system [53]. The authors of [54] proposed a unique version of an artificial bee colony model that can adaptively select an optimal search equation to estimate Turkey's energy consumption more accurately. The authors of [55] proposed a combined ant colony and genetic algorithm optimization model that powers an expert system with the ability to capture and simulate energy demand fluctuations under the influence of various factors. While these and similar studies have explored the application of biologically inspired algorithms in economics, few have compared the performance of different algorithms on a diverse set of economic tasks, as researchers often focus on a specific task at hand, treating the algorithm as a tool to obtain a desired outcome. In contrast, in this study, we analyze the appropriateness of biologically inspired algorithms for solving varying economic tasks compared to alternative autoML models (which are discussed next).

2.2. Automatic Machine Learning

With the increasing volume of data generated by individuals, organizations, and states, there is a growing need for automated solutions to help process, analyze, and extract insights from the data. Data-driven models based on ML models emerged as a possible solution to this need. However, the traditional ML process is time-consuming, requires substantial expertise, and can be error-prone. AutoML has materialized as an approach that automates many of the steps in the ML pipeline, including data preprocessing, feature engineering, model selection, and hyperparameter tuning, making most of the difficulties associated with using ML models irrelevant [56].

While ML models have demonstrated their impact in economic research, AutoML has yet to be fully adopted by economists [57–59]. In the context of our study, AutoML models can be roughly divided into two main groups: biologically inspired and non-biologically inspired models. Biologically inspired AutoML models are a specific instance of biologically inspired optimization algorithms that aims to find a suitable ML model for a given task. These models are commonly designed to mimic the search processes that occur naturally in nature or are utilized by various species. For instance, the Tree-based Pipeline Optimization Tool (TPOT) library used a genetic algorithm search process to find ML pipelines based on the popular Scikit-learn library [60]. On the other hand, the latter group of AutoML models includes a large number of methods. For example, the AutoSklearn library [61] also uses several search methods to build an ML pipeline on the Scikit-learn library, such as meta-learning over the dataset where the search starts from pipelines that performed well on similar datasets.

3. Methods and Materials

In this section, we formally outline the dataset gathering process followed by the models implemented on these datasets. Then, we describe the statistical analysis used on the obtained results from the experimental pipeline.

3.1. Datasets

We manually picked 50 datasets from Data World, focusing on different computational economic tasks. In order to focus only on datasets suitable for ML-based analysis, we included only datasets with either a time series or regression tasks. In addition, we excluded any dataset that had less than 300 instances (rows) and 5 features (columns) for a regression task and less than 50 instances and 20 features for the time series task. These sizes are typically considered the minimum amount of data to consider a task appropriate for ML-based analysis [62]. Each dataset is represented using a single table (matrix), and any numerical column where the number of unique values was more than a quarter of the number of instances was removed in order to avoid uninformative features [63]. We ensured that the datasets were indeed economy-related by relying on three independent economic scholars who did not co-author this paper. (Each of the economists has a Ph.D. in economics and has published at least two manuscripts in journals classified under the economy categories according to Web Of Science in the last two years.) Only datasets marked as economy-related by all three of them were considered. The datasets roughly belong to three disciplines in economics: (1) the stock market, (2) social policy funding, and (3) goods consumption.

3.2. Models

We focused on six models divided into three groups: traditional, biologically inspired AutoML, and other AutoML models. For the first group, we used the linear regression model implemented using the least mean square algorithm [64] due to its popularity in the economic domain [65,66]. In addition, we used a minimal decision tree model [67] implemented using the branch-and-bound search approach and a black box model guessing strategy proposed by [68]. This model mimics the way an expert would propose a rule-based model. For the biologically inspired AutoML group, we included TPOT [69] and the GPlearn [70] models. The former uses a genetic algorithm to search for an optimal ML pipeline while the latter utilizes an evolutionary algorithm to search for equations that best describe the data. However, for the latter, we replaced the standard fundamental functions, such as addition and multiplication, with ML models. Lastly, the non-biologically inspired AutoML group included AutoSklearn [61] and AutoGluon [71]. Both libraries utilized multiple computational ideas to make the ML pipeline search process accurate and effective. These models were chosen due to their popularity in both academia and industry [72–74].

3.3. Evaluation Process

We implemented the following evaluation using the Python programming language [75] (version 3.7.5). The analysis was implemented in two phases: meta-dataset generation and analysis. During the first phase, for each dataset in the set of datasets, we computed the mean square error (MSE), mean absolute error (MAE), and coefficient of determination (R^2) for each model. In particular, we repeat this computation twice and generate three results as follows. For the first iteration, we used all the samples in the dataset to compute a "fitting" performance of the model for each metric. Next, we split the dataset into "training" and "testing" cohorts such that the testing cohort contained the last (if a time series, so temporally ordered) 20% of the samples, while the training cohorts contained the remaining samples. Each model was then trained on the training cohort and evaluated both on the training and testing cohorts, resulting in additional two outcomes for each metric. Specifically, for the training of the model, we aimed to minimize the MAE metric. In order to test the models "as is", no preprocessing or hyperparameter tuning was performed outside of the implementation of each tool. During the second phase, we analyzed the obtained meta-dataset statistically, comparing the performance of each modeling group to the other two groups, which were divided into the three metrics and training cohorts.

Importantly, due to the stochastic and iterative nature of AutoML models, the more computational time these models have, the more they are commonly producing ever-improving results. In addition, the minimal decision tree is extremely computationally expensive. Hence, in order to make this a fair comparison, all methods were limited to 30 min for each dataset, running on a high-quality machine which was dedicated to this task. (The machine had an Intel i7 10th generation 10700k CPU with 16 GB of memory and running on an Ubuntu 18.04 operation system.) The evaluated models and the associated datasets were examined one after the other in a completely random order.

4. Results

The results of the analysis were obtained and stored in nine meta-datasets (three cohorts multiplied by the three metrics) and provided as supplementary material. To summarize these results, we computed the mean $\pm$ standard deviation of the $n = 50$ datasets such that the best result of each algorithmic group was taken for each case, as presented in Table 1. Notably, since the MAE and MSE metrics were supported between $[0, \infty)$ and sensitive to the absolute values of the dataset, we normalized the MAE and MSE metrics to be the relative MAE and MSE of each method compared to the worst MAE and MSE scores of each dataset, respectively. We marked these metrics as R-MAE and R-MSE, respectively, ranging between 0 and 1. However, this resulted in larger values indicating worse models, which was inconsistent with the R^2 metric. To make the results easier to interpret, we report the $1 - x$ score, where x is the R-MAE (R-MSE). For each metric, we computed an Anova test with post hoc t-tests with Bonferroni correction. Our results show that the biologically inspired models outperformed the other examined model groups using either the R-MAE or R^2 metrics at $p < 0.01$ for all comparisons. However, for the R-MSE metric, the other AutoML group (i.e., non-biologically inspired) outperformed the other groups at $p < 0.05$. In addition, we computed the number of cases for the test cohort, where at least one AutoML method and all four AutoML methods outperformed the traditional methods, reaching 87% and 73%, respectively.

Table 1. Summary of the central results, divided between the fitting, training, and testing cohorts. The results are shown as the mean $\pm$ standard deviation of $n = 50$ datasets such that the best result of each group is taken for each case. The best result for each metric and test cohort is marked in bold.

Test Cohort	Algo-Group	R-MSE	R-MAE	R^2
Fitting	Traditional	0.32 ± 0.18	0.41 ± 0.15	0.86 ± 0.11
	BI-AutoML	0.93 ± 0.10	$\mathbf{0.89 \pm 0.07}$	0.96 ± 0.05
	AutoML	$\mathbf{0.95 \pm 0.06}$	0.85 ± 0.07	$\mathbf{0.97 \pm 0.08}$
Training	Traditional	0.27 ± 0.14	0.38 ± 0.13	0.89 ± 0.09
	BI-AutoML	0.82 ± 0.10	$\mathbf{0.84 \pm 0.09}$	$\mathbf{0.95 \pm 0.09}$
	AutoML	$\mathbf{0.87 \pm 0.06}$	0.78 ± 0.07	0.95 ± 0.10
Testing	Traditional	0.16 ± 0.04	0.13 ± 0.02	0.58 ± 0.34
	BI-AutoML	0.91 ± 0.05	$\mathbf{0.95 \pm 0.05}$	$\mathbf{0.72 \pm 0.28}$
	AutoML	$\mathbf{0.93 \pm 0.04}$	0.92 ± 0.03	0.70 ± 0.30

5. Discussion and Conclusions

In this study, we evaluated the performance of three groups of algorithms on a versatile set of 50 real-world economic datasets: traditional models, biologically inspired AutoML, and other AutoML models. Our findings suggest that the traditional models were outperformed by all AutoML models for 73% of the cases and in 87% of the cases by at least one of the AutoML models. This outcome agrees with a larger trend in which advanced ML models, such as ones obtained by AutoML models, typically outperform traditional models, such as linear regression and minimal decision tree [76–78]. Table 1 reveals that the biologically inspired AutoML models, on average and with statistical significance, outperformed the other AutoML models across three metrics: the MAE, MSE, and R^2. The slightly favorable results obtained by the biologically inspired AutoML models compared with the other AutoML models further support and strengthen the apparent connection between biology and economics [36,37].

Nonetheless, this study is not without limitations. First, in our analysis, we chose to primarily focus on regression tasks. Other types of ML-based tasks such as classification tasks are prominent in economics [79], and as such, the generalization of our results to such cases needs further investigation. Second, we did not perform any preprocessing or feature engineering on the datasets. Such processes, when performed correctly, can alter the results significantly. Third, as we evaluated two models from each group of models, a possible extension to our work is the consideration of additional models. Third, models are commonly measured by more metrics than their performance, such as computational time and stability. Especially in economic fields like trading, these properties of the data-driven algorithm are of great interest. Hence, further investigation of the biologically inspired autoML models' performance based on these metrics is promising for future work.

Our study sheds light on the performance of different ML models in economics and highlights the potential benefits of using AutoML models, specifically biologically inspired ones, for data-driven tasks in this realm. More specifically, our results further support the emerging established connection between biology and economics. Future studies could build upon our findings by testing additional models on larger and more diverse datasets and by exploring different preprocessing and feature engineering methods. Ultimately, we hope that by following the proposed results, more professionals from economics would consider adopting AutoML models in their research and practice, contributing to economic sustainability projects as well as economics as a whole.

Author Contributions: Conceptualization, methodology, software, formal analysis, investigation, resources, writing—original draft, visualization, supervision, and project administration, T.L.; software, formal analysis, data curation, and writing—review and editing, T.F.; investigation, validation, and writing—review and editing, A.Y.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the code and data used in this work are freely available at https://github.com/teddy4445/bio_inspired_automl_economy.

Acknowledgments: The authors wish to thank Labib Shami for his economics-related consulting.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ML	Machine learning
AI	Artificial intelligence
AutoML	Automatic machine learning

References

1. Coglianese, C. *Optimizing Regulation for an Optimizing Economy*; Institute for Law and Economics, University of Pennsylvania: Philadelphia, PA, USA, 2018; pp. 18–35.
2. Lee, D.; Saez, E. Optimal minimum wage policy in competitive labor markets. *J. Public Econ.* **2012**, *96*, 739–749. [CrossRef]
3. Brannlund, R.; Nordstrom, J. Carbon tax simulations using a household demand model. *Eur. Econ. Rev.* **2004**, *48*, 211–233. [CrossRef]
4. Shrestha, R.M.; Marpaung, C.O.P. Supply- and demand-side effects of carbon tax in the Indonesian power sector: An integrated resource planning analysis. *Energy Policy* **1999**, *27*, 185–194. [CrossRef]
5. Shrestha, R.M.; Shrestha, R.; Bhattacharya, S.C. Environmental and electricity planning implications of carbon tax and technological constraints in a developing country. *Energy Policy* **1998**, *26*, 527–533. [CrossRef]
6. Cohen, N.; Klenk, T.; Davidovitz, M.; Cardaun, S. Varieties of welfare markets from a street-level perspective: Comparing long-term care services in Germany and Israel. *Public Adm. Rev.* **2023**, *83*, 419–428. [CrossRef]
7. Willenborg, M. Empirical Analysis of the Economic Demand for Auditing in the Initial Public Offerings Market. *J. Account. Res.* **1999**, *37*, 225–238. [CrossRef]
8. Arrow, K.J. Statistics and Economic Policy. *Econometrica* **1957**, *25*, 523–531. [CrossRef]
9. Shami, L.; Cohen, G.; Akirav, O.; Herscovici, A.; Yehuda, L.; Barel-Shaked, S. Informal Self-Employment within the Non-Observed Economy of Israel. *Int. J. Entrep. Small Bus.* 2021, *forthcoming*.
10. Dybka, P.; Kowalczuk, M.; Olesiński, B.; Torój, A.; Rozkrut, M. Currency demand and MIMIC models: Towards a structured hybrid method of measuring the shadow economy. *Int. Tax Public Financ.* **2019**, *26*, 4–40. [CrossRef]

11. Ha, L.T.; Dung, H.P.; Thanh, T.T. Economic complexity and shadow economy: A multi-dimensional analysis. *Econ. Anal. Policy* **2021**, *72*, 408–422. [CrossRef]
12. Guinee, J.B.; Heijungs, R.; Huppes, G. Economic allocation: Examples and derived decision tree. *Int. J. Life Cycle Assess.* **2004**, *9*, 23–33. [CrossRef]
13. Drabiková, E.; Škrabulakova, E.F. Decision trees—A powerful tool in mathematical and economic modeling. In Proceedings of the 2017 18th International Carpathian Control Conference (ICCC), Sinaia, Romania, 28–31 May 2017; pp. 34–39.
14. Molnau, W.E.; Montgomery, D.C.; Runger, G.C. Statistically constrained economic design of the multivariate exponentially weighted moving average control chart. *Qual. Reliab. Eng. Int.* **2001**, *17*, 39–49. [CrossRef]
15. Lin, S.N.; Chou, C.Y.; Wang, S.L.; Liu, H.R. Economic design of autoregressive moving average control chart using genetic algorithms. *Expert Syst. Appl.* **2012**, *39*, 1793–1798. [CrossRef]
16. Wang, P.; Fan, E.; Wang, P. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognit. Lett.* **2021**, *141*, 61–67. [CrossRef]
17. Tanwar, S.; Bhatia, Q.; Patel, P.; Kumari, A.; Singh, P.K.; Hong, W.C. Machine Learning Adoption in Blockchain-Based Smart Applications: The Challenges, and a Way Forward. *IEEE Access* **2020**, *8*, 474–488. [CrossRef]
18. Rana, R.; Staron, M.; Hansson, J.; Nilsson, M.; Meding, W. A framework for adoption of machine learning in industry for software defect prediction. In Proceedings of the 2014 9th International Conference on Software Engineering and Applications (ICSOFT-EA), Vienna, Austria, 29–31 August 2014; pp. 383–392.
19. Simon-Keren, L.; Liberzon, A.; Lazebnik, T. A computational framework for physics-informed symbolic regression with straightforward integration of domain knowledge. *Sci. Rep.* **2023**, *13*, 1249. [CrossRef]
20. Rokach, L. Decision forest: Twenty years of research. *Inf. Fusion* **2016**, *27*, 111–125. [CrossRef]
21. Lazebnik, T.; Bahouth, Z.; Bunimovich-Mendrazitsky, S.; Halachmi, S. Predicting acute kidney injury following open partial nephrectomy treatment using SAT-pruned explainable machine learning model. *BMC Med. Inform. Decis. Mak.* **2022**, *22*, 133. [CrossRef]
22. Gogas, P.; Papadimitriou, T.; Sofianos, E. Forecasting unemployment in the Euro area with machine learning. *J. Forecast.* **2022**, *41*, 551–566. [CrossRef]
23. Rosenfeld, A.; Cohen, M.; Taylor, M.E.; Kraus, S. Leveraging human knowledge in tabular reinforcement learning: A study of human subjects. *Knowl. Eng. Rev.* **2018**, *33*, e14. [CrossRef]
24. Athey, S.; Imbens, G.W. Machine Learning Methods That Economists Should Know About. *Annu. Rev. Econ.* **2019**, *11*, 685–725. [CrossRef]
25. Yoon, J. Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. *Comput. Econ.* **2021**, *57*, 247–265. [CrossRef]
26. Nosratabadi, S.; Mosavi, A.; Duan, P.; Ghamisi, P.; Filip, F.; Band, S.S.; Reuter, U.; Gama, J.; Gandomi, A. Data Science in Economics: Comprehensive Review of Advanced Machine Learning and Deep Learning Methods. *Mathematics* **2020**, *8*, 1799. [CrossRef]
27. Litleskare, S.; Wuyts, W. Planning Reclamation, Diagnosis and Reuse in Norwegian Timber Construction with Circular Economy Investment and Operating Costs for Information. *Sustainability* **2023**, *15*, 10225. [CrossRef]
28. Kairiss, A.; Geipele, I.; Olevska-Kairisa, I. Sustainability of Cultural Heritage-Related Projects: Use of Socio-Economic Indicators in Latvia. *Sustainability* **2023**, *15*, 10109. [CrossRef]
29. He, X.; Zhao, K.; Chu, X. AutoML: A Survey of the State-of-the-Art. *Knowl.-Based Syst.* **2021**, *212*, 106622. [CrossRef]
30. Chen, Y.W.; Song, Q.; Hu, X. Techniques for automated machine learning. *ACM SIGKDD Explor. Newsl.* **2021**, *22*, 35–50. [CrossRef]
31. Real, E.; Liang, C.; So, D.; Le, Q. Automl-zero: Evolving machine learning algorithms from scratch. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 8007–8019.
32. Wang, W.; Xu, W.; Yao, X.; Wang, H. Application of Data-driven Method for Automatic Machine Learning in Economic Research. In Proceedings of the 2022 21st International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Chizhou, China, 14–18 October 2022; pp. 42–45.
33. Chen, W.; Zhang, L. An automated machine learning approach for earthquake casualty rate and economic loss prediction. *Reliab. Eng. Syst. Saf.* **2022**, *225*, 108645. [CrossRef]
34. Agrapetidou, A.; Charonyktakis, P.; Gogas, P.; Papadimitriou, T.; Tsamardinos, I. An AutoML application to forecasting bank failures. *Appl. Econ. Lett.* **2021**, *28*, 5–9. [CrossRef]
35. Singpai, B.; Wu, D. Using a DEA–AutoML Approach to Track SDG Achievements. *Sustainability* **2020**, *12*, 10124. [CrossRef]
36. Szekely, P.; Sheftel, H.; Mayo, A.; Alon, U. Evolutionary Tradeoffs between Economy and Effectiveness in Biological Homeostasis Systems. *PLoS Comput. Biol.* **2013**, *9*, e1003163. [CrossRef] [PubMed]
37. Lacoste, D.; Rivoire, O.; Tourigny, D.S. Cell behavior in the face of uncertainty. *arXiv* **2023**, arXiv:2304.13733.
38. dos Santos Coelho, L.; Mariani, V.C. Use of chaotic sequences in a biologically inspired algorithm for engineering design optimization. *Expert Syst. Appl.* **2008**, *34*, 1905–1913. [CrossRef]
39. Zheng, C.; Sicker, D.C. A Survey on Biologically Inspired Algorithms for Computer Networking. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 1160–1191. [CrossRef]
40. Ferri, G.; Caselli, E.; Mattoli, V.; Mondini, A.; Mazzolai, B.; Dario, P. SPIRAL: A novel biologically-inspired algorithm for gas/odor source localization in an indoor environment with no strong airflow. *Robot. Auton. Syst.* **2009**, *57*, 393–402. [CrossRef]

41. Hassanat, A.B.A.; Alkafaween, E. On Enhancing Genetic Algorithms Using New Crossovers. *Int. J. Comput. Appl. Technol.* **2017**, *55*, 202–212. [CrossRef]
42. Poli, R.; Kennedy, J.; Blackwell, T. Particle swarm optimization. *Swarm Intell.* **2007**, *1*, 33–57. [CrossRef]
43. Dasgupta, D. Advances in artificial immune systems. *IEEE Comput. Intell. Mag.* **2006**, *1*, 40–49. [CrossRef]
44. Tang, W.J.; Wu, Q.H. Biologically inspired optimization: A review. *Trans. Inst. Meas. Control* **2009**, *31*, 495–515. [CrossRef]
45. Kobayashi, M.H. On a biologically inspired topology optimization method. *Commun. Nonlinear Sci. Numer. Simul.* **2010**, *15*, 787–802. [CrossRef]
46. Li, Y.W.; Wüst, T.; Landau, D.P. Biologically Inspired Surface Physics: The HP Protein Model. In *Nanophenomena at Surfaces: Fundamentals of Exotic Condensed Matter Properties*; Michailov, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 169–183.
47. Macy, M. Natural Selection and Social Learning in Prisoner's Dilemma: Coadaptation with Genetic Algorithms and Artificial Neural Networks. *Sociol. Methods Res.* **1996**, *25*, 103–137. [CrossRef]
48. Bellaachia, A.; Bari, A. Flock by Leader: A Novel Machine Learning Biologically Inspired Clustering Algorithm. In *Advances in Swarm Intelligence*; Tan, Y., Shi, Y., Ji, Z., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 117–126.
49. Li, Z.; Liu, J. A multi-agent genetic algorithm for community detection in complex networks. *Phys. A Stat. Mech. Its Appl.* **2016**, *449*, 336–347. [CrossRef]
50. Chung, H.; Shin, K.S. Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction. *Sustainability* **2018**, *10*, 3765. [CrossRef]
51. Skolpadungket, P.; Dahal, K.; Harnpornchai, N. Portfolio optimization using multi-obj ective genetic algorithms. In Proceedings of the 2007 IEEE Congress on Evolutionary Computation, Singapore, 25–28 September 2007; pp. 516–523.
52. Hong, W.C.; Dong, Y.; Chen, L.Y.; Wei, S.Y. SVR with hybrid chaotic genetic algorithms for tourism demand forecasting. *Appl. Soft Comput.* **2011**, *11*, 1881–1890. [CrossRef]
53. Kuo, R.J.; Chen, C.H.; Hwang, Y.C. An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network. *Fuzzy Sets Syst.* **2001**, *118*, 21–45. [CrossRef]
54. Özdemir, D.; Dörterler, S.; Aydın, D. A new modified artificial bee colony algorithm for energy demand forecasting problem. *Neural Comput. Appl.* **2022**, *34*, 17455–17471. [CrossRef]
55. Ghanbari, A.; Kazemi, S.M.R.; Mehmanpazir, F.; Nakhostin, M.M. A Cooperative Ant Colony Optimization-Genetic Algorithm approach for construction of energy demand forecasting knowledge-based expert systems. *Knowl.-Based Syst.* **2013**, *39*, 194–206. [CrossRef]
56. Yao, Q.; Wang, M.; Chen, Y.; Dai, W.; Li, Y.F.; Tu, W.W.; Yang, Q.; Yu, Y. Taking Human out of Learning Applications: A Survey on Automated Machine Learning. *arXiv* **2019**, arXiv:1810.13306.
57. Shami, L.; Lazebnik, T. Implementing Machine Learning Methods in Estimating the Size of the Non-observed Economy. *Computational Economy* **2023**. [CrossRef]
58. Rakshit, S.; Clement, N.; Vajjhala, N.R. Exploratory Review of Applications of Machine Learning in Finance Sector. In *Advances in Data Science and Management*; Borah, S., Mishra, S.K., Mishra, B.K., Balas, V.E., Polkowski, Z., Eds.; Springer: Singapore, 2022.
59. Warin, T.; Stojkov, A. Machine Learning in Finance: A Metadata-Based Systematic Review of the Literature. *J. Risk Financ. Manag.* **2021**, *14*, 302. [CrossRef]
60. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
61. Feurer, M.; Klevin, A.; Eggensperger, K.; Springenberg, J.T.; Blum, M.; Hutter, F. Auto-sklearn: Efficient and Robust Automated Machine Learning. In *Automated Machine Learning*; Springer: Cham, Switzerland, 2019.
62. Feldman, D.; Schmidt, M.; Sohler, C. Turning big data into tiny data: Constant-size coresets for k-means, PCA, and projective clustering. *SIAM J. Comput.* **2020**, *49*, 601–657. [CrossRef]
63. Chandrashekar, G.; Sahin, F. A survey on feature selection methods. *Comput. Electr. Eng.* **2014**, *40*, 16–28. [CrossRef]
64. Transtrum, M.K.; Sethna, J.P. Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization. *arXiv* **2012**, arXiv:1201.5885.
65. Adrian, T.; Crump, R.K.; Moench, E. Pricing the term structure with linear regressions. *J. Financ. Econ.* **2013**, *110*, 110–138. [CrossRef]
66. Panwar, B.; Dhuriya, G.; Johri, P.; Singh Yadav, S.; Gaur, N. Stock Market Prediction Using Linear Regression and SVM. In Proceedings of the 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 4–5 March 2021; pp. 629–631.
67. Carrizosa, E.; Molero-Río, C.; Romero-Morales, D. Mathematical optimization in classification and regression trees. *TOP* **2021**, *29*, 5–33. [CrossRef]
68. McTavish, H.; Zhong, C.; Achermann, R.; Karimalis, I.; Chen, J.; Rudin, C.; Seltzer, M. Fast Sparse Decision Tree Optimization via Reference Ensembles. In Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2022; pp. 9604–9613.
69. Olson, R.S.; Moore, J.H. TPOT: A tree-based pipeline optimization tool for automating machine learning. In Proceedings of the Workshop on Automatic Machine Learning, New York, NY, USA, 24 June 2016; pp. 66–74.
70. Fortin, F.A.; De Rainville, F.M.; Gardner, M.A.; Parizeau, M.; Gagné, C. DEAP: Evolutionary Algorithms Made Easy. *J. Mach. Learn. Res.* **2012**, *13*, 2171–2175.

71. Erickson, N.; Mueller, J.; Shirkov, A.; Zhang, H.; Larroy, P.; Li, M.; Smola, A. AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data. *arXiv* **2020**, arXiv:2003.06505.
72. Nagarajah, T.; Poravi, G. A Review on Automated Machine Learning (AutoML) Systems. In Proceedings of the 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Pune, India, 29–31 March 2019; pp. 1–6.
73. Karmaker, S.K.; Hassan, M.; Smith, M.J.; Xu, L.; Zhai, C.; Veeramachaneni, K. AutoML to Date and Beyond: Challenges and Opportunities. *ACM Comput. Surv.* **2021**, *54*, 1–36. [CrossRef]
74. Lazebnik, T.; Somech, A.; Weinberg, A.I. SubStrat: A Subset-Based Optimization Strategy for Faster AutoML. *Proc. VLDB Endow.* **2022**, *16*, 772–780. [CrossRef]
75. Srinath, K.R. Python—The Fastest Growing Programming Language. *Int. Res. J. Eng. Technol.* **2017**, *4*, 354–357.
76. Vadyala, S.R.; Betgeri, S.N.; Matthews, J.C.; Matthews, E. A review of physics-based machine learning in civil engineering. *Results Eng.* **2022**, *13*, 100316. [CrossRef]
77. Shehab, M.; Abualigah, L.; Shambour, Q.; Abu-Hashem, M.A.; Shambour, M.K.Y.; Alsalibi, A.I.; Gandomi, A.H. Machine learning in medical applications: A review of state-of-the-art methods. *Comput. Biol. Med.* **2022**, *145*, 105458. [CrossRef] [PubMed]
78. Thai, H.T. Machine learning for structural engineering: A state-of-the-art review. *Structures* **2022**, *38*, 448–491. [CrossRef]
79. Li, H.; Sun, J.; Wu, J. Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods. *Expert Syst. Appl.* **2010**, *37*, 5895–5904. [CrossRef]