Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Systematic Literature Review of Machine Learning Algorithms for Predicting Customer Churn in the context of HRMS Software Providers

Soniya Lalwani[1]
[1]*Department of Computer Science and Information Technology, Parul University Gujarat, India*
[1]*sonia.ccca@gmail.com*

ARTICLE INFO

Received: 30 Dec 2024

Revised: 19 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

Introduction: The Churn means Attrition. In the world of competition and customer needs, the superior quality service, to predict customer churn, has become a pressing issue for the HRMS Vendors or service providers. Machine Learning is capable to solve this challenging problem. The customer churn can be predicted, and proactively, HRMS vendors can apply a strategy to retain customers and manage the customer relationship.

Objectives: The objective of this paper is to meticulously analyse and summarize the previous studies conducted for customer churn prediction. Furthermore, this paper will serve as a valuable resource for future research in the context of HRMS vendors or service providers.

Methods: This paper focuses on machine learning algorithms like Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting, Neural Network, etc. The paper highlights the accuracy of the outperformed models. This paper includes the previous studies between 2021 and 2025. In total, 30 papers were reviewed meticulously. To provide more clarity, all the previous studies are summarized in a table format. The paper was searched on Google with the keywords 'Customer Churn' AND 'Machine Learning'. The Duplicates were removed, and the previous studies were analyzed.

Results: The findings showcase that various machine learning models like RF, DT, Logistic Regression, Gradient Boosting are applied to predict churn. Furthermore, by this analysis it is clearly stated that hybrid or ensemble models work best for customer churn model prediction. This study guides on selecting the best model as per the relevant industry.

Conclusions: This comprehensive survey has meticulously analysed the implementation of machine learning algorithms to predict customer churn and highlighted its importance to fully transform the business to build client retention strategies. In this paper, different machine models are examined, and for each previous study, the machine learning models used & their results are discussed.

Keywords: Customer Churn, HRMS, HRMS Software Providers, Churn Prediction, Machine Learning, Customer Retention

INTRODUCTION

Churn means the state when the customer is not using services or products, or you can say the customer is not satisfied and is not using our service or product, which means the customer is churned. If there is customer churn constantly taking place in the company, then it directly impacts the business and its objectives. How is the business impacted? There are a few components in the HRMS Software licensing, like: First time license activation cost, one time installation and setup cost, Annual Maintenance Subscription cost, Employee upgrade cost, Third party application integration cost, customization and customization maintenance cost. Traditionally, HRMS Vendors were

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

using the application logs to check module usage. HRMS Vendors collect general feedback from customers to check the happiness score.

Here comes Machine Learning, which plays a crucial role in solving customer churn-related problems. Using machine learning, we get predictive analysis of churn. Machine learning algorithms can understand the hidden patterns and complex relationships. Ultimately, using Machine learning models, HRMS software providers will have automation in churn prediction, which in turn can be used to retain those customers who are on the verge of churn by providing discounts, gifts, etc.

OBJECTIVES

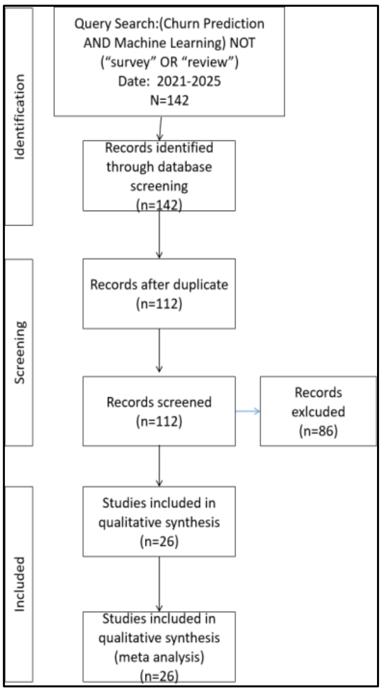
The objective and scope of this Review Paper are to systematically and meticulously study and analyse existing studies on ML Models for churn prediction. In this paper, there will be an exploration of ML Models, Data Pre-processing Techniques Used, Hyper parameter Tuning, Handling of Class Imbalance issues, Results, and Research Gaps. Special attention will be given to these studies from the context of HRMS Software Providers. This paper will serve as a foundation for future research work in the context of HRMS software provided by companies.

The remainder of this paper is as follows: In Section 2, there is a research problem. In Section 3 Methodology of the Literature Review is described. In sections 4 & 5, Classification of Machine Learning Models and Literature review and Tabular projection of previous studies. In section 6, Research Gap. In Section 7, Future Work. In the last section, the Conclusion is discussed.

METHODS

A systematic literature search was conducted on various academic journals, publications including Shodh Ganga - Inflibnet, Bulletin of Electrical Engineering and Informatics, IEEE Explore, MDPI, International Journal for Research in Applied Science and Engineering Technology (IJAI), Technoscientifica, DRPress, BC Publication, Proceeding International Conference on Science and Engineering, and many more.

The research was restricted to the publishing years between 2021 to 2025. We have tried to keep the study as possible. As illustrated in Figure 1, in the initial search we found 142 articles. We screened the identified articles and removed the duplicates. After removing duplicates, 112 unique articles were left. From 112 unique articles we have excluded 86 articles to keep the study domain specific and most relevant. Figure 1. PRISMA Flowchart is drawn below.



Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Table 1: Classification Models used to predict customer churn are as follows:

Model Name	Abbreviation	Explanation
Logistic Regression	LR	It predicts outcomes in 0 or 1 based on features
Decision Tree	DT	It is like Branches of Tree predicts "Yes" and "No" based on feature values
Random Forest	RF	It combines Multiple Decision Tree Predictions and considers the average of all the Decision Tree predictions
Support Vector Machine	SVM	It separates the classes by the best hyperplane
K-Nearest Neighbors	KNN	The majority of labels are classified based on nearest neighbours
Naive Bayes	NB	It is a probability model based on Bayes' Theorem
Gradient Boosting	GB	In this technique the model is updated sequentially by correcting its previous iteration error.
XG Boost	XGB	It is an improved version of Gradient Boosting
Ada Boost	Ada Boost	In this technique, the miss classified instances are re-weighted to form a strong classifier by combining weak learners
K-Means Clustering	K-Means Clustering	In each cluster the variance is minimized in K clusters. Data is grouped into K Clusters.
Hierarchical Clustering	Hierarchical Clustering	The nested clusters are built by merging and splitting on distance metrics
Principal Component Analysis	PCA	In this technique, the features of the dataset are reduced and transformed into principal components
Artificial Neural Networks	ANN	It functions like the human brain, and through the neuron layer, complex patterns are learned
Convolution Neural Networks	CNN	The data is processed in the form of grid-like data
Recurrent Neural Network	RNN	Through feedback loops, the memory is maintained, and it is designed for sequence data.

Literature review and Tabular projection of previous studies:

In this section overview of previous studies in customer churn prediction is presented. Industries like Telecom, Retail, E-Commerce, Banking, Hotel, etc are considered.

Nagaraju, J. (2024). [24], The author focused on the insurance industry, they collected an insurance dataset. The author used Meta-heuristic models like Binary Golden Eagle Optimizer (BGEO) for feature selection. Furthermore, the author has used Techniques such as Firefly Enhanced, Particle Swarm Optimization (PSO), Genetic Algorithm, and Ant Colony Optimization (ACO) to select the relevant features. The author has used base models like Artificial Neural Networks (ANN), Decision Trees (DT), Logistic Regression (LR), Random Forest (RF), and Support Vector Machines (SVM). The author used hybrid Models like Firefly Enhanced AdaBoost Ensemble Classifier (FFE-AdaBoost), Customized Extreme Learning Machine (CELM), and Blended Logistic Regression Decision Tree (BLRDT). They also used Deep Learning Techniques like Recurrent Neural Network (RNN), Long Short-Term

Memory (LSTM), and Gated Recurrent Unit (GRU). As a result, FFE-AdaBoost ensemble model achieved 97.12% accuracy, and the CELM model achieved 96.4% accuracy.

S Brinthakumari, Priyanka Dhiraj Sananse, Punam Bagul. (2023). [19], Focused on the telecommunication industry. The author proposed an mSVM (modified SVM) approach. They used a BigML Dataset containing 5000 records. To preprocess the data, they used Textual analysis, Elimination of stop words, stemming (using Porter's method), Index term extraction, Data filtering, Lexical analysis, Standardization, extraction, and selection of characteristics techniques. The authors used machine learning algorithms like Modified Support Vector Machine (mSVM), Naïve Bayes (NB), K-Nearest Neighbor (KNN), Decision Tree (DT) with J48 classifiers, Logistic Regression (LR), Random Forest (RF), Multilayer Perceptron (MLP) neural network, Bayesian Network. The author revealed that mSVM outperformed and gave 95% accuracy as compared to other machine learning algorithms. In this study author discussed that the combination of NLP and ML called mSVM is best on the telecommunication dataset.

Saha, S., Saha, C., Haque, Md. M., Alam, Md. G. R., & Talukder, A. (2024). [23], The author focused on the telecommunications industry. They used a deep learning model to improve the customer churn prediction accuracy, and they named it the ChurnNet model. They used the IBM Telco Dataset with 7043 records and 21 features, the Churn-in Telecom Dataset with 3333 records and 21 features and the Churn-data-UCI Dataset with 5000 records and 20 features. Handled Missing Value with Null value treatment, for categorical features, label encoding and one-hot encoding techniques were used. The authors used SMOTE, SMOTETomek, and SMOTEEN techniques for handling class imbalance issues. They used machine learning models like Churnet, including 1D CNN, Residual Block, Squeeze and Excitation Blocks, Spatial Attention Modules and found that Churnet achieved the best result on the Churn-Data-UCI Dataset, 97.52% accuracy.

Table 2: A Summary of Previous Studies to Predict Customer Churn

Year	Industry	Author	ML Model used	Result
2021	Telecom	S, M. (2021). [1]	Adaptive Logit Boost, Multivariate Logistic Regression, Distance Measure with Gaming Theory	ALB 92.46%
2021	Telecom	P, S. (2021). [2]	Improvised XGBoost, Modified Random Forest (MRF)	Improvised XGBoost 99.54%
2021	Motor Insurance	Das, D. (2021).[3]	Hybrid GWO-KELM Model. Combines Grey Wolf Optimization (GWO) for feature selection with Kernel Extreme Learning Machine (KELM) for classification	KELM 95% Accuracy
2022	Banking	Elyusufi, Y., & Kbir, M. A. (2022). [4]	LR,SVM,KNN,NB,RF,Ensemble Model	RF 89.07% Accuracy
2022	Insurance	Kingawa, E. D., & Hailu, T. T. (2022). [5]	K-means++, LR, NB, DT, RF, SVM, DNN	DNN 98.81% Accuracy
2022	Telcom	Vasudevan, M., Narayanan, R. S., Nakeeb, S. F., & Abhishek, A. (2022). [6]	XG Boost, RF, NB	Xgboost 80.30% accuracy
2022	Telcom	Priyanka, P. G., & Rahaman, Mr. Sk. A. (2022). [7]	LR, DT Regressor, RF Regressor, XGboost Regressor,	XG Boost Regressor 91.20%

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

2022	Telcom	Mahalekshmi, A., & Chellam, G. H. (2022). [8]	LR, R, XGBoost, Gradient Boos Adaboost, ANN, CNN, S-tacked Autoencoders	CNN performed best with 84.12% accuracy
2022	SaaS	Çalli, L., & Kasim, S. (2022). [9]	DT, LR, NB, KNN, RF, NN	RF 78.27% accuracy
2022	Telcom	National School of Applied Sciences, & Loukili, M. (2022). [10]	LR, SM, KN N, DT	SVM 96.92% accuracy
2022	Telcom	Sana, J. K., Abedin, M. Z., Rahman, M. S., & Rahman, M. S. (2022). [11]	KNN, NB, LR, RF, DT, GB, FFN,RNN were used	FNN combined with WOE performed best
2022	Banking	Leena Mandurkar, Sawali Khanke, Rani Khandaskar, & Anmol Ukey. (2022). [12]	KNN, DT, RF Classifier, XGBoost, SVM, LR	Not discussed which Model performed the best
2022	Telcom	Krishnan R, Cv Krishnaveni, & Av Krishna Prasad. (2022). [13]	LR, Bayesian Model, Random Forest, Gradient Boosting, Decision Tree	XG Boost 93%
2022	Telecom	Vani, K. (2022). [14]	DT with Genetic Algorithm and Hill Climbing	75% Accuracy
2022	Telecom	Ramesh, P. (2022). [15]	RF, ANN, ANN with 2 hidden layers, ANN with 4 hidden layers, Structure Optimized Simulated Annealing ANN (SA-ANN), Structure Optimized Population-Based Incremental Learning ANN (PBIL-ANN), Structure Optimized Hybrid SA-PBIL ANN, PBIL-Wide and Deep Learning Model (PBIL-WDLM)	PBIL-ANN 4.46% higher accuracy than ANN 2 layers
2023	Banking	Dang Tran, H., Le, N., & Nguyen, V.-H. (2023). [16]	KNN, LR, DT, RF, SVM	RF 97%
2023	Telcom	Maan, J., & Maan, H. (2023). [17]	LR, DT, RF, XG Boost	XG Boost 96.2%
2023	Telecom, finance, banking, and e-commerce	American University in the Emirates, Dubai, UAE, Abualkishik, A. Z., American University in the Emirates, Dubai, UAE, R., Towson University,	Stacked Deep Learning with Wind Driven Optimization, CCPBI-TAMO, PCPM, LSTM-SAE, SVM	Stacked Deep Learning with Wind Driven Optimization 98.60%

		Towson University, Maryland's University, USA, & Thompson, W. (2023). [26]		
2023	Banking	Chen, P., Liu, N., & Wang, B. (2023). [18]	LR, SV M, GBDT, Adaboost	Adaboost 71.8% recall
2023	telecom	S Brinthakumari, Priyanka Dhiraj Sananse, Punam Bagul. (2023). [19]	Modified Support Vector Machine (mSVM) Naïve Bayes (NB) K-Nearest Neighbor (KNN) Decision Tree (DT) with J48 classifiers Logistic Regression (LR) Random Forest (RF) Multilayer Perceptron (MLP) neural network Bayesian Network	mSVM 95%
2023	Banking	Zhao, S. (2023).[20]	RF & DT	RF 91%
2023	Telecom, OTT Platforms, Automotive Subscription, Airlines	Zhang, B. (2023). [21]	LR, GB, NN, DT, RF, KNN, GNB, Light GBM, XGBoost	LR 79.6%
2023	Telecom	Hangarge, P., Jadhav, G., Janagave, V., Kadam, S., & Pise, Prof. P. S. (2023). [22]	SVM, RF, XGBoost, Ridge Classifier, NN	SVM Best
2024	Telecom	Saha, S., Saha, C., Haque, Md. M., Alam, Md. G. R., & Talukder, A. (2024). [23]	Churnet including 1D CNN, Residual Block, Squeeze and Excitation Blocks, Spatial Attention Modules.	Churn achieved the best result on the Churn-Data-UCI Dataset 97.52%
2024	Insurance	Nagaraju, J. (2024). [33]	Base model: Artificial Neural Networks (ANN), Decision Trees (DT), Logistic Regression (LR), Random Forest (RF), Support Vector Machines (SVM).Hybrid Models: Firefly Enhanced AdaBoost Ensemble	FFE-AdaBoost ensemble model achieved 97.12% accuracy for churn prediction. The

Journal of Information Systems Engineering and Management
2025, 10(42s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

			Classifier (FFE-AdaBoost), Customized Extreme Learning Machine (CELM), Blended Logistic Regression Decision Tree (BLRDT).Deep Learning Techniques: Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU).	CELM model demonstrated 96.4% accuracy, outperforming other individual models in its category.
2023	General	Bhoite, S. N., Gadekar, V. D., Kapadnis, S. V., & Ghuge, P. R. (2023). [25]	R F, Light GBM, SVM, LR	RF 86.01%

RESULTS

Based on the literature review it is identified that no focus is provided to HRMS Vendors. Majorly, the churn prediction study is conducted for telecommunication, banking, retail, e-commerce, SaaS industry. Further to this, almost all the studies have considered the open dataset available on the internet or websites like Kaggle. There is a need to experiment on real real-world dataset. In some research papers class imbalance issue is not discussed, or the evaluation metric is different, which makes the simple comparison task the most difficult one.

DISCUSSION

Specific to HRMS Software provider companies model evaluation process is much needed. There should be a benchmark for the HRMS Vendors' specific datasets. Furthermore, deeper learning models must be tailored specifically to the HRMS software providers' patterns. We have meticulously analysed the literature from 2021 to 2025 on machine learning techniques used to predict customer churn in the context of HRMS. We have analysed that various machine learning models like RF, DT, Logistic Regression, and Gradient Boosting are applied to predict churn. From this survey we have analysed that hybrid or ensemble models work best for customer churn model prediction. The analysis clearly shows that there is a research gap in the context of HRMS Vendors and there is a strong urge to further research on HRMS and to develop a predictive model to help maintain the profitable business of HRMS Vendors. Lastly, we can say that there is room for further research, specifically customizing it to the HRMS specific domain, to aid the HRMS Vendors to effectively manage their customer relationship and maintain their profitability from the business.

REFERENCES

[1] S, M. (2021). Customer Churn Prediction in Telemarketing Industry Using Machine Learning Technique. University. Retrieved from https://shodhganga.inflibnet.ac.in/handle/10603/397866
[2] P, S. (2021). Customer churn prediction in the telecom domain using Machine Learning Algorithms. University. Retrieved from https://shodhganga.inflibnet.ac.in/handle/10603/471774
[3] Das, D. (2021). Machine learning model for customer attrition prediction in the motor insurance sector. University. Retrieved from https://shodhganga.inflibnet.ac.in/handle/10603/444251
[4] Elyusufi, Y., & Kbir, M. A. (2022). Churn Prediction Analysis by Combining Machine Learning Algorithms and Best Features Exploration. International Journal of Advanced Computer Science and Applications, 13(7). https://doi.org/10.14569/IJACSA.2022.0130773
[5] Kingawa, E. D., & Hailu, T. T. (2022). Customer Churn Prediction Using Machine Learning Techniques: The Case of Lion Insurance. Asian Journal of Basic Science & Research, 04(04), 60–73. https://doi.org/10.38177/AJBSR.2022.4407
[6] Vasudevan, M., Narayanan, R. S., Nakeeb, S. F., & Abhishek, A. (2022). Customer churn analysis using XGBoosted decision trees. Indonesian Journal of Electrical Engineering and Computer Science, 25(1), 488. https://doi.org/10.11591/ijeecs.v25.i1.pp488-495

[7] Priyanka, P. G., & Rahaman, Mr. Sk. A. (2022). Customer Churn Prediction in Telecom Industry Using Regression Algorithms. International Journal of Innovative Research in Computer Science & Technology, 54–57. https://doi.org/10.55524/ijircst.2022.10.3.10
[8] Mahalekshmi, A., & Chellam, G. H. (2022). Analysis of customer churn prediction using machine learning and deep learning algorithms. International Journal of Health Sciences. https://doi.org/10.53730/ijhs.v6nS1.7861
[9] Çalli, L., & Kasim, S. (2022). Using Machine Learning Algorithms to Analyze Customer Churn in the Software as a Service (SaaS) Industry. Academic Platform Journal of Engineering and Smart Systems, 10(3), 115–123. https://doi.org/10.21541/apjess.1139862
[10] National School of Applied Sciences, & Loukili, M. (2022). Supervised Learning Algorithms for Predicting Customer Churn with Hyperparameter Optimization. International Journal of Advances in Soft Computing and Its Applications, 14(3), 50–63. https://doi.org/10.15849/IJASCA.221128.04
[11] Sana, J. K., Abedin, M. Z., Rahman, M. S., & Rahman, M. S. (2022). A novel customer churn prediction model for the telecommunication industry using data transformation methods and feature selection. PLOS ONE, 17(12), e0278095. https://doi.org/10.1371/journal.pone.0278095
[12] Leena Mandurkar, Sawali Khanke, Rani Khandaskar, & Anmol Ukey. (2022). Churn Prediction using Various Machine Learning Algorithms. International Journal of Advanced Research in Science, Communication and Technology, 429–435. https://doi.org/10.48175/IJARSCT-7707
[13] Krishnan R, Cv Krishnaveni, & Av Krishna Prasad. (2022). Telecom Churn Prediction using Machine Learning. World Journal of Advanced Engineering Technology and Sciences, 7(2), 087–096. https://doi.org/10.30574/wjaets.2022.7.2.0130
[14] Vani, K. (2022). Mining Mobile Data for Telecom Churn Management. University. Retrieved from https://shodhganga.inflibnet.ac.in/handle/10603/419430
[15] Ramesh, P. (2022). Hybrid machine learning techniques for predicting customer churn. University. Retrieved from https://shodhganga.inflibnet.ac.in/handle/10603/534326
[16] Dang Tran, H., Le, N., & Nguyen, V.-H. (2023). Customer Churn Prediction in the Banking Sector Using Machine Learning-Based Classification Models. Interdisciplinary Journal of Information, Knowledge, and Management, 18, 087–105. https://doi.org/10.28945/5086
[17] Maan, J., & Maan, H. (2023). Customer Churn Prediction Model using Explainable Machine Learning. https://doi.org/10.48550/ARXIV.2303.00960
[18] Chen, P., Liu, N., & Wang, B. (2023). Evaluation of Customer Behaviour with Machine Learning for Churn Prediction: The Case of Bank Customer Churn in Europe. Proceedings of the International Conference on Financial Innovation, FinTech and Information Technology, FFIT 2022, October 28-30, 2022, Shenzhen, China. Proceedings of the International Conference on Financial Innovation, FinTech and Information Technology, FFIT 2022, October 28-30, 2022, Shenzhen, China, Shenzhen, People's Republic of China. https://doi.org/10.4108/eai.28-10-2022.2328450
[19] S Brinthakumari, Priyanka Dhiraj Sananse, Punam Bagul. (2023). mSVM: An Approach for Customer Churn Prediction using modified Support Vector Machine and various Machine Learning Techniques. Proceedings International Conference on Science and Engineering, 11(1), 1158–1169. https://doi.org/10.52783/cienceng.v11i1.258
[20] Zhao, S. (2023). Customer Churn Prediction Based on the Decision Tree and Random Forest Model. BCP Business & Management, 44, 339–344. https://doi.org/10.54691/bcpbm.v44i.4840
[21] Zhang, B. (2023). Customer Churn in Subscription Business Model—Predictive Analytics on Customer Churn. BCP Business & Management, 44, 870–876. https://doi.org/10.54691/bcpbm.v44i.4971
[22] Hangarge, P., Jadhav, G., Janagave, V., Kadam, S., & Pise, Prof. P. S. (2023). Customer Churn Prediction Using Machine Learning Algorithms. International Journal for Research in Applied Science and Engineering Technology, 11(4), 2374–2376. https://doi.org/10.22214/ijraset.2023.50646
[23] Saha, S., Saha, C., Haque, Md. M., Alam, Md. G. R., & Talukder, A. (2024). ChurnNet: Deep Learning Enhanced Customer Churn Prediction in Telecommunication Industry. IEEE Access, 12, 4471–4484. https://doi.org/10.1109/ACCESS.2024.3349950
[24] Nagaraju, J. (2024). Customer Churn Prediction in Insurance Sector using Machine Learning and Optimization Techniques. University. Retrieved from https://shodhganga.inflibnet.ac.in/handle/10603/558758

[25] Bhoite, S. N., Gadekar, V. D., Kapadnis, S. V., & Ghuge, P. R. (2023). Churn Prediction using Machine Learning Models. International Journal for Research in Applied Science and Engineering Technology, 11(5), 2233–2236. https://doi.org/10.22214/ijraset.2023.52101

[26] American University in the Emirates, Dubai, UAE, Abualkishik, A. Z., American University in the Emirates, Dubai, UAE, R., Towson University, Towson University, Maryland's University, USA, & Thompson, W. (2023). Intelligent Model for Customer Churn Prediction using Deep Learning Optimization Algorithms. Journal of Intelligent Systems and Internet of Things, 8(1), 43–54. https://doi.org/10.54216/JISIoT.080104