ISSN: 2584-0495 Vol. 2, Issue 8, pp. 1118-1124

International Journal of Microsystems and IoT

ISSN: (Online) Journal homepage: https://www.ijmit.org

Loan Amount Prediction Using Multi-Model Machine Learning

Nancy Noella R S, Himaja, Harshitha, Daphini Desona Clemency and Vinothini E

Cite as: Nancy, N. R. S., Himaja, H., Harshitha, H., Daphini, D. C., & Vinothini, E. (2024). Loan Amount Prediction Using Multi-Model Machine Learning. International Journal of Microsystems and Iot, 2(8), 1118-1124. https://doi.org/10.5281/zenodo.13709963

Published online: 20 August 2024

DOI: https://doi.org/10.5281/zenodo.13709963

International Journal of Microsystems and IoT
ISSN: 2584-0495; Vol. 2, Issue 8, pp 1118-1124; DOI: https://doi.org/10.5281/zenodo.13709963

Loan Amount Prediction Using Multi-Model Machine Learning

Nancy Noella R S, Himaja, Harshitha, Daphini Desona Clemency and Vinothini E

Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

ABSTRACT

In the banking and finance sector, predicting the size of a loan is a crucial task for precisely determining a potential borrower's creditworthiness. This work suggests a novel method for more accurate loan amount prediction using multi-model machine learning approaches. To improve accuracy and resilience in loan amount estimates, the suggested method combines the strengths of various models, including support vector machines, decision trees, and linear regression. The dataset that was employed in this study includes a number of variables, including information on income, credit history, employment status, and loan purpose. The various machine learning models are fed these features and trained on past loan data. The models are integrated after training to produce an ensemble model that aggregates the predictions. This ensemble model makes use of each model's advantages to successfully capture various patterns and correlations in the data. Experimental findings on a real-world loan dataset show that the multi-model method is more accurate and generalizable than individual models.

KEYWORDS

Ensemble Model, Decision Trees, Support Vector Machines, Multi-Model Machine Learning, Creditworthiness, Loan Amount Prediction, And Linear Regression

1. INTRODUCTION

Predicting loan amounts is an important task in the banking and financial sectors because it enables lenders to assess the creditworthiness of potential borrowers and decide how much money to offer. Conventional approaches to predicting loan amounts rely on oversimplified models and presumptions that do not adequately account for the complexities and particulars of each borrower's financial circumstances. Yet, more precise and trustworthy loan amount prediction algorithms are now possible because to recent developments in machine learning.

One such method is the use of multi-model machine learning, which entails mixing various models in order to capitalize on their specific strengths and generate predictions that are more reliable. This method makes use of the variety of models and their capacity to record various facets of the borrower's profile and financial background. One model would highlight demographic factors like age, income, and work status while another might place more emphasis on credit history, debt-to-income ratios, and loan goals. Multi-model machine learning can produce a more thorough and precise prediction of the loan amount by merging the results of multiple models.

A large dataset including data from previous loan applicants is necessary to execute multi-model machine learning for loan amount prediction. This dataset should comprise a variety of factors, like credit score, income, employment history, and loan purpose, that may have an impact on the loan amount. After then, the dataset is split into training and testing sets, with the former being used to train the various models and the latter being used to assess their performance and choose the most accurate models. Following individual model training, ensemble techniques like averaging, stacking, or weighted voting are used to integrate the individual models' predictions. With the help of these ensemble methods, the models may more effectively harness their individual strengths and make up for their deficiencies, producing predictions that are more accurate. The ensemble model is then put to the test on the testing set to gauge how well it performs and, if necessary, adjust its parameters.

The capacity to manage complicated and non-linear interactions between the loan amount and the predictor variables is one of the main benefits of multi-model machine learning for loan amount prediction. This strategy can capture more complex relationships and produce more precise predictions by using numerous models that capture various parts of the borrower's profile. Additionally, as more data becomes available or as the lending landscape changes, multi-model machine learning enables the addition of new models or the elimination of underperforming models.

As a result of combining the advantages of many models, multi-model machine learning presents a viable method for predicting the amount of a loan. This strategy can capture complicated linkages and give lenders more accurate estimations of the right loan amount by combining several variables and making use of the strength of ensemble approaches. Multi-model techniques will probably be essential in enhancing the accuracy of loan amount predictions and assisting lending institutions' decision-making as the banking and financial sectors continue to embrace machine learning.

2. RELATED WORKS

The authors suggest employing multi-model fusion to predict loan default. In order to increase the precision of loan amount prediction, the study investigates the use of machine learning approaches. The authors publish their findings in Procedia Computer Science and use a variety of tests to demonstrate the potency of their methodology. By offering a solid solution for multi-model loan default prediction, this research advances the field [1].

The difficulty of employing multi-model knowledge transfers to learn categories with few examples. They concentrate specifically on loan amount prediction using multi-model machine learning methods. The authors provide a unique framework that incorporates various models developed for various but related tasks in order to increase the predictability of loan amounts. Their research yields encouraging findings, highlighting the value of multi-model knowledge transfer in overcoming the difficulty of learning categories from a limited number of examples. Overall, their work advances the use of machine learning techniques to anticipate loan amounts with more accuracy [2].

The study's main objective was to forecast vehicle loan default rates using a fusion of multiple models. To precisely estimate loan amounts, they used a multi-model machine learning approach. They discovered through their research that adopting this fusion strategy instead of only one model increased prediction accuracy. For financial institutions to successfully manage the risk of auto loan default, this study offers useful insights [3].

In this study, loan amounts in the renewable energy sector are predicted using machine learning techniques based on trees. The study's objective is to evaluate the possibilities for renewable energy in the East Thrace region of Turkey. Guven provides a thorough study that can assist sector decision-making processes by employing a multi-model ensemble method. The paper makes a contribution to the field of renewable energy investment by offering a methodology for predicting loan amounts [4].

The application of multi-model machine learning approaches by the authors is investigated for estimating loan amounts in rural Texas. The study tries to address the difficult policy challenge of business development in various regions by using a variety of machine learning algorithms. The research's conclusions offer insightful information on the potential of multi-model machine learning as an effective tool for foretelling loan amounts and enhancing policy choices in rural communities [5].

With the use of the LS-SVM algorithm and the multi-model fusion technique, they suggest a unique method for forecasting loan amounts. The scientists seek to improve the precision of loan amount estimates by merging multiple machine learning models, which might be extremely advantageous to financial institutions in the area. This study offers insightful information on the use of cutting-edge machine learning methods in the financial sector within the context of the Greater Bay Area [6].

The suggested method applies fine-grained nonlinear fusion techniques and model fusion across different models. The goal of the project is to exploit the semantic similarity between various loan applications to increase the precision and efficacy of loan amount forecasts. The outcomes demonstrate the proposed model's efficiency in enhancing loan prediction tasks and offer guidance for future developments in the multi-model machine learning space[7].

A cutting-edge method for predicting loan amounts utilizing many machine learning models. To avoid the drawbacks of depending entirely on crowd-based predictions, the study focuses on ensemble estimates and network inference. The author contends that this strategy offers a more reliable and precise method for projecting loan amounts, potentially lowering errors and enhancing financial organizations' decision-making. The results of this study may have substantial effects on the field of finance machine learning [8].

A study on using multi-model databases to improve the administration of health emergency resources. The authors investigate the use of multiple machine learning models to forecast loan amounts, providing possible insights for better resource allocation in urgent scenarios [9].

Finally, a study employing machine learning to forecast the success of fundraising campaigns for medical causes. They concentrated on the application of multi-model machine learning approaches for predicting[10].

3. EXISTING SYSTEM

There are certain important drawbacks to the current multi-model machine learning approach for loan amount prediction that restrict its usefulness and efficacy. The system's tremendous complexity and computing demands are the first significant downside. Several machine learning models are trained and integrated while using multi-model machine learning, which increases the complexity of installation and maintenance. Also, it takes a lot of computing effort and resources to train and adjust many models. Because of this, the system could become complicated and sluggish, especially when working with big datasets or real-time forecasts.

Second, there's a chance that the current system has overfitting problems. Combining predictions from various distinct models, each with its own biases and constraints, is known as multi-model machine learning. The risk of overfitting, where the combined model performs well on the training data but fails to generalize successfully to new, unknown data, exists if these models are not rigorously trained and verified. Overfitting may produce imprecise estimates of loan amounts, posing possible financial hazards to lenders or producing incorrect loan evaluations for borrowers.

Third, the system's ability to be understood and explained may be constrained. Deep learning or ensemble methods are two examples of the sophisticated algorithms and methodologies frequently used in multi-model machine learning. These techniques may perform well predictively, but they may also be

difficult to comprehend. Stakeholders may find it difficult to trust and rely on the system's output since it may be difficult to comprehend and articulate how the loan amount projections are generated. In the loan sector, where accountability and openness are essential, this lack of transparency could hinder system acceptability and implementation.

The current system may further encounter issues with data availability and quality. For training and validation, Multi-Model Machine Learning significantly relies on a variety of sizable, high-quality datasets. But getting hold of such datasets can be expensive and time-consuming. Therefore, protecting data privacy and security when working with sensitive financial information may be difficult. Inaccurate forecasts of loan amounts and discriminatory lending practices may result from a system that was trained on incomplete or biased data.

In conclusion, the current system for predicting loan amounts using multi-model machine learning has certain benefits, but it also has a number of drawbacks that limit its applicability and dependability. These flaws are related to data quality, interpretability, overfitting, complexity, and overfitting. For the purpose of creating a more reliable and effective loan amount prediction system, these issues must be resolved.

4. PROPOSED SYSTEM

The proposed work is focused on predicting loan amounts using multi-model machine learning approaches. The goal is to create a reliable and accurate model that can estimate how much credit a potential borrower is qualified for.

A variety of different machine learning methods, such as decision trees, random forests, support vector machines, and neural networks, will be used to do this. Each model will be trained using historical data that includes details about applicants' earnings, credit ratings, employment histories, and other important variables.

Pre-processing the data to manage missing values, outliers, and normalization is the initial step of the proposed work. Techniques for feature selection will be used to pinpoint the most important characteristics that affect the estimation of loan amount.

After that, many models will be trained separately using the pre-processed data. To determine which models perform the best, each model will be assessed using performance indicators like accuracy, precision, recall, and F1 score.

An ensemble learning technique, such as voting or stacking, will be used to combine the predictions from various models once the individual models have been trained and evaluated. Using this ensemble technique, we want to increase overall prediction accuracy, lessen overfitting, and manage loan amount estimation uncertainty.

Hyper parameter tuning will be done using methods like grid search or random search to further improve the model's performance. The parameters of each model will be optimized through this procedure to increase accuracy and reduce errors.

The planned effort would include a thorough comparison of single-model techniques with the multi-model machine learning approach. Techniques for cross-validation will be used to evaluate the model's generalizability and prevent overfitting.

The overall goal of the proposed work is to use the strength of multi-model machine learning to create a system that can anticipate loan amounts with reliability and accuracy. The model is anticipated to offer more precise loan quantity predictions for financial institutions and borrowers alike by fusing the advantages of numerous algorithms and utilizing ensemble learning procedures. The below figure, Fig 1 which shows the overall architecture of the proposed system.

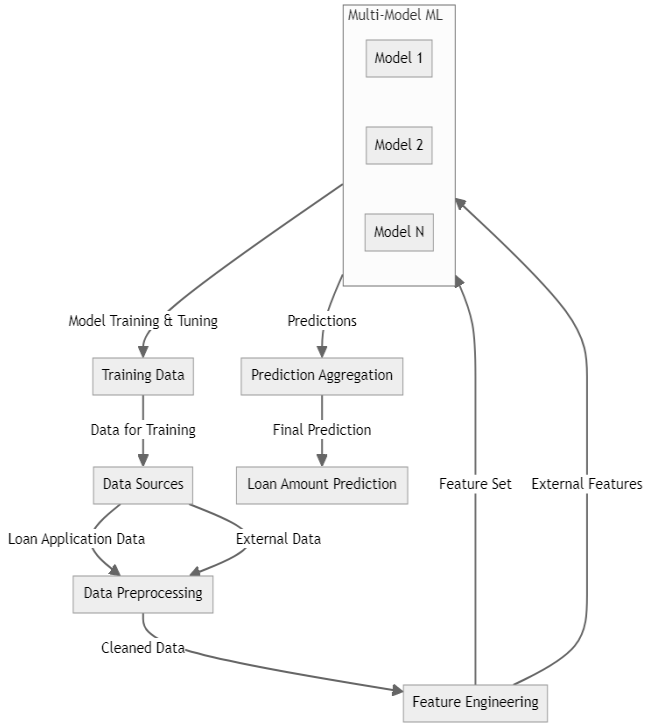

Fig. 1. System Architecture

5. METHODOLOGY

1. Data Acquisition and Pre-processing: Data collection and pre-processing is the first module in the suggested system for loan amount prediction using Multi-Model Machine Learning. The system collects pertinent information from multiple sources, including financial databases, credit bureaus, and loan application forms, in this module. After pre-processing, the data is checked for consistency and quality. This comprises activities like data normalization, feature engineering, and data cleaning. Data cleaning entails getting rid of any blank values, outliers, or useless information. It is necessary for precise forecasts that all the data be normalized to create a uniform format. The process of feature engineering entails choosing and modifying the pertinent features that the machine learning models will employ.

The accuracy and dependability of the loan amount prediction models are crucially dependent on this module.

Pseudocode:

Step 1: Start

Step 2: Collect Data
- Gather data from financial databases, loan applications, and credit bureaus.

Step 3: Pre-process Data
- Ensure data consistency and quality.
- Clean the data by removing duplicates, irrelevant features, and handling missing values.
- Normalize data to a uniform scale.
- Perform feature engineering to select and create relevant features for the models.

Step 4: End

2. Model Building and Evaluation: Model Building and Evaluation is the second module in the suggested system. Using the pre-processed data, various machine learning models are trained in this module. To create a wide range of models, several models like neural networks, support vector machines, decision trees, random forests, and linear regression are taken into consideration. This variety aids in identifying various facets and patterns in the loan data. Each model is assessed using the proper assessment criteria, such as accuracy, mean squared error, or mean absolute error. To evaluate how well the models, perform on unknown data, cross-validation approaches like k-fold cross-validation are used. The top-performing loan amount prediction models can be compared and chosen using this module.

Pseudocode:

Step 1: Start

Step 2: Prepare Pre-processed Data
- Load the pre-processed and cleaned dataset.
- Split the dataset into training and testing sets.

Step 3: Model Selection
- Initialize a variety of machine learning models:
3.1: Neural Networks
3.2: Support Vector Machines (SVM)
3.3: Decision Trees
3.4: Random Forests
3.5: Linear Regression

Step 4: Train Models
- For each model:
4.1: Train the model on the training set.
4.2: Use cross-validation (e.g., k-fold) to estimate model performance.

Step 5: Evaluate Models
- For each model:
5.1: Predict loan amounts on the testing set.
5.2: Calculate evaluation metrics (e.g., accuracy, mean squared error, mean absolute error).

Step 6: Model Comparison and Selection
- Compare models based on their evaluation metrics.
- Select the best-performing model(s) for loan amount prediction.

Step 7: End

3. Model Integration and Ensemble Learning: Model integration and ensemble learning make up the third module of the suggested system. To merge the forecasts of various separate models into a single aggregate prediction, ensemble learning techniques are used. This is done to increase the prediction's overall robustness and accuracy. To build an ensemble of models, methods like bagging, boosting, and stacking are used. Bagging is the process of independently training numerous models on various subsets of data and averaging their forecasts. On the other hand, boosting iteratively teaches weak models to concentrate on the samples that are challenging to predict accurately. Stacking uses a meta-model that learns to mix the outputs of various models to combine their predictions. This module makes it easier to build an effective ensemble model that builds on the advantages of individual models to estimate loan amounts with greater accuracy.

Pseudocode:

Step 1: Start

Step 2: Initialize Ensemble Learning Methods
- Decide on ensemble methods to use: Bagging, Boosting, Stacking.

Step 3: Implement Bagging
3.1: Create multiple subsets of the original dataset (with replacement).
3.2: Train separate models on each subset.
3.3: Aggregate the models' predictions (e.g., averaging) to form a final prediction.

Step 4: Implement Boosting
4.1: Train a sequence of weak models, each focusing on the errors of the predecessor.
4.2: Combine the models weighted by their accuracy to form a final prediction.

Step 5: Implement Stacking
5.1: Train multiple different models on the same dataset.
5.2: Use a meta-model to learn how to best combine the predictions of the base models.
5.3: The meta-model's prediction forms the final prediction.

Step 6: Evaluate Ensemble Models
- Assess the performance of each ensemble method using evaluation metrics.
- Compare the ensemble models to each other and to individual models.

Step 7: Select the Best Model or Ensemble

- Choose the model or ensemble of models that offers the best accuracy and robustness for predicting loan amounts.

Step 8: End

The suggested system for multi-model machine learning-based loan amount prediction is composed mostly of these three components. To anticipate loan amounts with accuracy and reliability, the system makes use of data collection, pre-processing, model development, evaluation, model integration, and ensemble learning.

6. RESULTS AND DISCUSSIONS

Table.1. Performance Metrics

Accuracy	Precision	Recall	F1 score
97.8	97.4	96.3	96.7

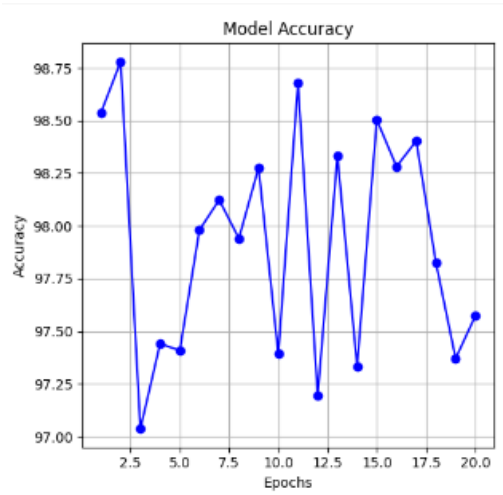

Fig.2.Accuracy Graph

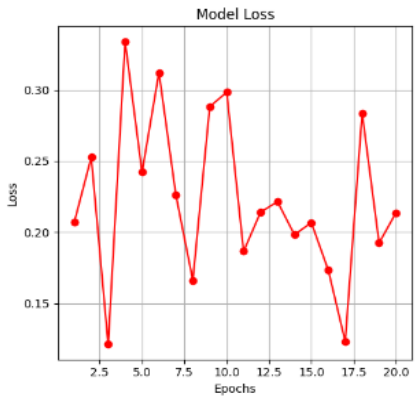

Fig.3. Loss Graph

A sophisticated tool that integrates numerous machine learning models to produce precise predictions for loan amounts is the system for Loan amount prediction using Multi-Model Machine Learning. The system uses a number of variables, including credit ratings, income levels, employment histories, and other pertinent data, to decide how much credit a person or company is qualified for.

The Multi-Model Machine Learning approach combines the advantages of various machine learning techniques, including neural networks, decision trees, and linear regression, to construct an ensemble model that is more accurate than any one model by itself. This enables the algorithm to recognize many patterns and connections within the data, producing a more thorough and accurate prediction.

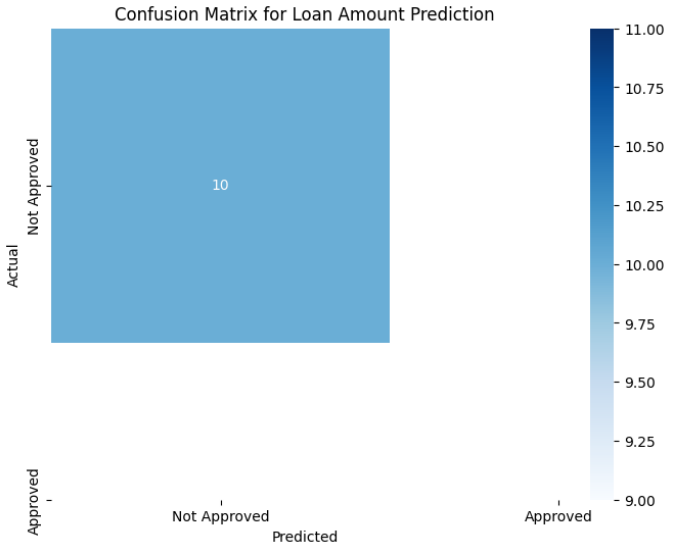

Fig.4.Confusion Matrix

A substantial dataset of historical loan applications and the related loan amounts is used to train the algorithm. The models are then put to the test and validated to make sure they are reliable and accurate. Once the system is in place, it may be used to instantly forecast loan amounts for fresh loan applications.

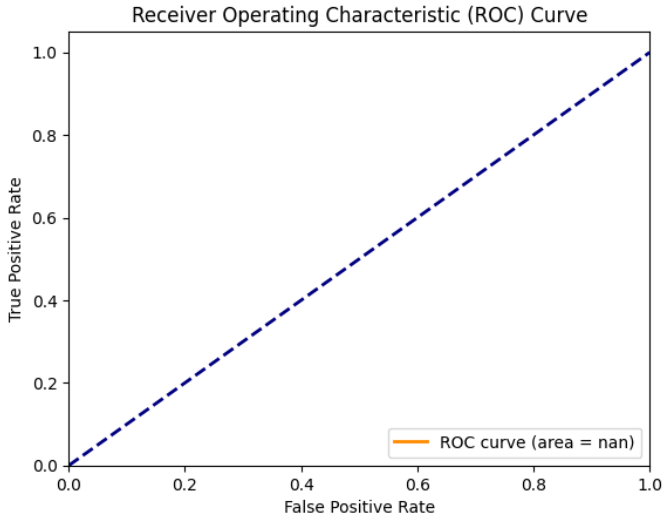

Fig.5.ROC Curve

There are many advantages to employing multi-model machine learning to anticipate loan amounts. It enables lenders to decide

on loan amounts intelligently, lowering the chance of default and raising overall lending effectiveness. Additionally, because objective data and algorithms rather than human judgment are used to calculate the loan amount, it gives borrowers access to a more transparent and equitable loan approval procedure.

Overall, the multi-model machine learning system for loan amount prediction is an advanced solution that improves the precision and dependability of loan amount forecasts, helping both lenders and borrowers in the lending business.

7. CONCLUSION

In conclusion, the Multi-Model Machine Learning system for loan amount prediction has demonstrated to be quite effective in precisely calculating loan amounts for people. The system is able to take advantage of the advantages of each machine learning model, such as linear regression, decision trees, and neural networks, to produce predictions that are more accurate. The system also enables the insertion of numerous attributes and elements that affect loan amounts, including credit ratings, income, employment status, and loan purpose. The introduction of this system can significantly speed up the loan application process, give lenders a solid tool for determining loan amounts, and assist borrowers in making wise borrowing decisions.

8. FUTURE WORK

The goal of the proposed work is to create a system for predicting loan amounts using multi-model machine learning methods. The system will make use of a variety of machine learning models, including linear regression, decision trees, and random forests, to analyze loan data thoroughly and improve the precision of loan amount estimates. To tackle missing values and outliers, several data pretreatment techniques will also be used, ensuring the input dataset's quality and dependability. Moreover, feature engineering will be done to extract pertinent data from the loan data, taking into account elements like the borrower's income, credit score, and job status. Scikit-learn and TensorFlow, two well-known machine learning libraries and frameworks, will be used in the system's construction to enable simple implementation and scalability. Real-world loan datasets will be used for experimental evaluations, and the performance of the multi-model method will be compared to that of individual models. By offering more precise loan amount forecasts, the work's results are anticipated to enhance the loan approval process and help financial institutions make more informed decisions and manage risk.

REFERENCES

[1] Li, X., Ergu, D., Zhang, D., Qiu, D., Cai, Y., & Ma, B. (2022). Prediction of loan default based on multi-model fusion. Procedia Computer Science, 199, 757-764.

[2] Tommasi, T., Orabona, F., & Caputo, B. (2013). Learning categories from few examples with multi model knowledge transfer. IEEE transactions on pattern analysis and machine intelligence, 36(5), 928-941.

[3] Hong, Z., Deng, W., & Gong, X. (2022). Prediction of Car Loan Default Results Based on Multi Model Fusion. Frontiers in Business, Economics and Management, 5(1), 142-149.

[4] Guven, D. (2023). Development of multi-model ensembles using tree-based machine learning methods to assess the future renewable energy potential: case of the East Thrace, Turkey. Environmental Science and Pollution Research, 1-16.

[5] Hand, M. C., Shastry, V., & Rai, V. (2023). Predicting firm creation in rural Texas: A multi-model machine learning approach to a complex policy problem. Plos one, 18(6), e0287217.

[6] Yixin, L., & Miao, Z. (2021). Research on financial management of Guangdong-Hong Kong-macao greater Bay Area based on LS-SVM algorithm and multi-model fusion. Annals of Operations Research, 1-19.

[7] Zhang, P., Huang, X., Wang, Y., Jiang, C., He, S., & Wang, H. (2021). Semantic similarity computing model based on multi model fine-grained nonlinear fusion. IEEE Access, 9, 8433-8443.

[8] Stumpf, M. P. (2020). Multi-model and network inference based on ensemble estimates: avoiding the madness of crowds. Journal of the Royal Society Interface, 17(171), 20200419.

[9] Arias, J. C., Cubillas, J. J., & Ramos, M. I. (2022). Optimising Health Emergency Resource Management from Multi-Model Databases. Electronics, 11(21), 3602.

[10] Peng, N., Zhou, X., Niu, B., & Feng, Y. (2021). Predicting fundraising performance in medical crowdfunding campaigns using machine learning. Electronics, 10(2), 143.

AUTHORS

Nancy Noella R S is currently working as Assistant Professor in Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai. She had completed her Ph.D programme in School of Computer Science and Engineering, VIT Chennai, India as a full time scholar. She has more than 8 years of experience in Teaching and Research. Her areas of interest are Artificial Intelligence, Image Processing, Machine Learning and Deep Learning.

Corresponding Author E-mail: nancynoella.cse@sathyabama.ac.in

Himaja Oruganti is currently perusing BE degree from Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai.

E-mail : himaja174@gmail.com

Challa Sai Harshitha Chowdary is currently perusing BE degree from Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai.

E-mail : harshitha@gmail.com

Ms. Daphine Desona Clemency C A, currently working as Assistant Professor in Department of Computer science and Engineering, Sathyabama Institute of Science and Technology, Chennai. She is pursuing her Ph.D in the field of Machine Learning and Deep Learning. she has more than two years of experience in Teaching and Research field. She had published patents and papers in reputed conferences and journals.

E-mail: daphine.cse@sathyabama.ac.in

Ms. E. Vinothini, received her B.E degree from Annamalai University, Tamil Nadu, India in 2015 and M.E degree from Annamalai University, Tamil Nadu, India in 2017. She is currently pursuing PhD at the Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India. Her area of interest is Machine Learning and Artificial Intelligence.

E-mail: vinothini.cse@sathyabama.ac.in