Journal of Information Systems Engineering and Management
2025, 10(9s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Optimized Fused Regression Model for Regression Algorithms

Swati Gupta[1(0009-0000-4890-9879)], Bal Kishan[2(0009-0001-3694-9627)]
[1]*Research Scholar, Department of Computer Science, Maharshi Dayanand University, Rohtak, Haryana, India*
[2]*Assistant Professor, Department of Computer Science, Maharshi Dayanand University, Rohtak, Haryana, India*
Corresponding author: Swati Gupta (swati.rs20.dcsa@mdurohtak.ac.in)

ARTICLE INFO

Received: 03 Nov 2024

Revised: 24 Dec 2024

Accepted: 06 Jan 2025

ABSTRACT

One or more independent variables are compared to a dependent variable using regression analysis. Prediction and inference are its key goals. This strategy helps identify data patterns and trends to estimate constant outputs from variables. This research examines how Gradient Adaptive Moment Estimation Optimiser and ensemble multiple linear regression may improve regression task prediction. OFRM efficacy is assessed using six datasets from distinct sectors. six datasets from different domains were utilised to test OFRM. Test it against five regression models. Apply strict criteria and test OFRM extensively to establish its impact on anticipated accuracy, robustness, and generalisability. OFRM dominates individual regression on all datasets. This research shows OFRM's performance in regression scenarios to advance ensemble learning. This paper emphasises the necessity to combine optimisation and ensemble techniques to enhance regression models for real-world applications. Regression model performance on NFT datasets was evaluated using MSE, RMSE, MAE, and R^3 measures. OGFR predicts accurately with the lowest MSE (1.04), RMSE (2.21), and MAE (1.29). Best fit is achieved with a R^2 value of 0.85, accounting for 85% of sample variance. By outperforming DNR and KNN with R^2 of 0.55 and MSE of 1.91, OGFR is the top model for NFT dataset predictions with a R^2 of 0.75 and MSE of 1.24.

Keywords: Regression algorithms, Bayesian regression, Sparse regression, Ensemble methods, multifaceted approach, Linear Regression, Optimized Fused Regression Model

INTRODUCTION

Fundamental to statistical modelling, regression analysis is necessary to determine the kind of interdependencies and project continuous outcomes. Over the years, numerous regression techniques have been developed, each with its own strengths and limitations. Traditional approaches, such as linear regression, provide interpretable models but may struggle to capture complex nonlinear relationships in data. Ensemble methods, which combine the predictions of multiple models, offer improved predictive performance but may lack interpretability. OFRM aims to leverage the iterative refinement of model parameters facilitated by gradient Adaptive Moment Estimation with the collective intelligence of ensemble multiple linear regression learning, resulting in enhanced predictive accuracy and model robustness. OFRM intends to tackle the issues using traditional regression techniques and provide new paths for predictive modeling in various domains by merging the two methodologies. The OFRM framework is thoroughly examined in the paper along with its theoretical foundations, pragmatic uses, and empirical evaluation. With an eye toward other well-known regression models including GBR, KNN, DNR, KR, and NR, this study mostly focuses on OFRM. Four performance criteria—the MSE, RMSE MAE, R-squared—as well as six datasets—credit card transactions, US accidents, California property values, gold price regression,bike sharing demand, and NFT dataset—evaluate the model. This work investigates OFRM's performance on many datasets and regression tasks by comparing its outcomes with those of solo regression models and more traditional ensemble methods. This aims to show, by means of thorough testing and meticulous analysis, that OFRM is superior than its rivals in respect to interpretability, generalizability, and forecast accuracy. This also looks at OFRM's inner workings, illustrating how gradient adaptive moment estimation optimization combined with ensemble methods enhances regression outcomes. We will next review some of the more pragmatic features, likely applications, and future OFRM research orientations. The main objective of this work is to enhance regression analysis as a whole by means of a new framework that, taken combined, raise predictive modelling to unprecedented levels and provide the foundation for next research and development in this field. Give academics and business leaders a powerful tool to handle challenging regression projects and receive insightful analysis of data with OFRM. Section 2 deals with traditional research. Section 3 details the Optimizing Fused Regressors concept. Section 4 will next go over the dataset and pre-processing techniques. Section 5 covers the OFRM procedure, setup, and results. Section 6 provides performance review and repercussions; Section 7 explores possible future scope and conclusions.

RELATED WORK

The literature presented covers in great detail regression modelling and its uses in many different disciplines. Yang and Li looked at resident travel patterns using a multivariate logistic regression model. Examining the components influencing travel decisions and patterns can help one to gain understanding of transportation planning and urban mobility [1]. Madhu Kumar et al. built a regression model for short-term load forecasting on university campuses. This research most likely looks at methods to properly forecast power demand in order to better control energy and distribute resources [2]. Asghar et al. discussed "RECLAIM," a demand-side management system using renewable energy and ML. This study likely proposes strategies to optimize energy consumption and integrate renewable sources into the power grid [3]. Sun et al. focused on estimating human body orientation using radar-based techniques and hierarchical regression models. The study likely explores applications in healthcare or security[4]. Yi et al. looked at phase identification in networks that distribute low voltage. It is very probable that this study will help enhance the efficiency and dependability of power distribution systems [5]. For multicollinear predictors and multivariate response data, Yu et al. suggested using partial least squares regression trees. For complicated datasets with associated variables, this study probably provides improvements in regression modelling methods [6]. Parametric software effort estimate was introduced by Nhung et al. using multiple linear regressions and optimizing correction factors. Accurate resource allocation and project planning are anticipated to be helped by this study, which presumably tackles issues in software project management [7]. Connectivity in penalized regression-based linear multivariate processes was the primary area of study for Antonacci et al. The research likely contributes to network analysis or signal processing fields, offering insights into complex system dynamics [8]. Brzyski et al. proposed matrix variate regression for estimating brain connectivity associated with clinical outcomes. This research likely contributes to neuroimaging or medical diagnostics, aiding in understanding brain function and disease progression [9]. Kashima et al. introduced a federated learning approach using linear regression for server aggregation. This research likely addresses privacy concerns in distributed machine learning systems while enabling collaborative model training [10]. He et al. presented expected regression with errors-in-variables, likely contributing to advancements in statistical modelling methods for skewed or asymmetric data distributions [11]. Using precise linear regression equations as its foundation, Chen et al. presented a data-driven approach to power flow. In order to improve grid stability and management, this study probably tackles problems with analysing and optimizing power systems [12]. Contributing to breakthroughs in computer vision and motion analysis applications, Gu et al. proposed bias-compensated integral regression for human posture estimation [13]. A framework for semi-supervised contrastive regression was introduced by Ge et al. for the purpose of mapping forest inventories using data collected by many satellite sensors. Environmental monitoring and resource management are two areas that might benefit from this study's findings [14]. Recent developments in regression-related machine learning algorithms were likely aided by the multi-kernel learning support vector regression ensemble approach with AdaBoost developed by Xie et al. [15]. An improved method for quickly computing k-fold cross-validation and excellent values for the regularization parameters in ridge regression was proposed by Liland et al. Methods for evaluating and selecting models are expected to benefit from this study's findings [16]. Research on sustainable mobility and energy infrastructure planning was likely aided by ElGhanam et al.'s data-driven approach to EV power demand modelling using spatial regression [17]. These studies all highlight how regression modelling has developed and its potential. Research underlined the need of distributed computing methods for scalability [18]. Data fusion was shown to be helpful in assessing agricultural performance [19]; research demonstrated the integration of impedance-based approaches with regression models for plant health evaluation [20]. The findings emphasise the need of improving regression models for numerous sectors, including large-scale computing frameworks, environmental monitoring, and precision agriculture, including those related.

OPTIMIZED FUSED REGRESSION MODEL (OFRM)

OFRM is an optimal regression framework as it combines the advantages of gradient adaptive moment estimation optimizer and ensemble multiple linear regression methods to improve prediction accuracy in regression challenges. OFRM trains several linear regression models iteratively using Gradient Adaptive Moment Estimation Optimizer to maximize prediction error. We then combine the forecasts from these underlying regression models using weighted ensemble techniques. The weights are changed continuously in line with model performance. By fusing Adaptive Moment Estimation Optimizer with ensemble learning, OFRM achieves improved accuracy, robustness, and generalization capabilities compared to standalone regression models and traditional ensemble methods. This synergistic approach allows OFRM to effectively capture complex relationships in the data and produce more reliable predictions, making it a valuable tool for a wide range of regression applications shown in figure 1.

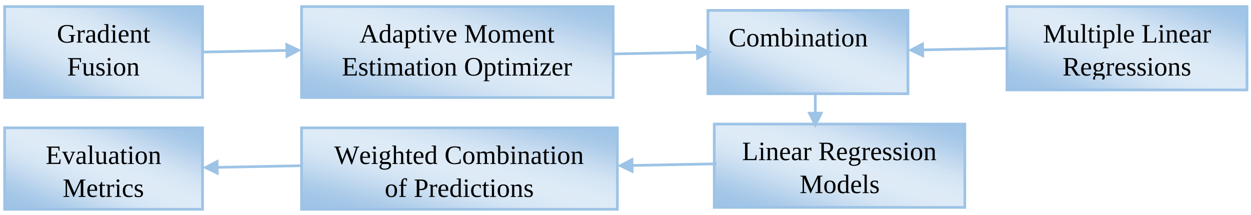

Figure 1. Layered framework of OGFR

SELECTION OF DATASET &PRE-PROCESSING

It includes data collection, analysis, and interpretation to achieve a goal.

Selection of datasets

This study focuses on textual dataset and is used for regression-based models during accuracy prediction. Total five data sets are selected, and their details are given in Table 1.

Table 1. Description of datasets

Dataset	Type of Dataset	Dataset Description
Transactions made by credit cards [21]	Textual	492 frauds out of 284,807 transactions, Time, Amount, Class, V1 to V28 of PCA
Gold price dataset [22]	Textual	Price, Date,Commodities, Economic indicators, and Forex rates
US Accidents [23]	Textual	Temperature (F), Humidity (%), Pressure (in), Wind_Chill (F), Visibility (mi), Target, Wind_Speed (mph)
California Housing Prices [24]	Textual	MedInc, AveRooms, HouseAge, AveBedrms, AveOccup, Population, Latitude, Longitude, Target
Bike Sharing Demand [25]	Textual	Season, working_day,holiday, weather, temp, atemp, windspeed, Target output, humidity
Empirical NFT Dataset [26]	Textual	NFT dataset generated with 1000 transactions

Preprocessing of Textual Datasets

- Tokenization: Firstly, Split the text into individual tokens to represent the input data in a format suitable for modelling.
- Lowercase: To ensure representation consistency, convert all text to lowercase and reduce the vocabulary size.
- Removal of stop words: Then, eliminate common words that have little semantic meaning and may introduce noise into the data.
- Stemming or Lemmatization: Then, to normalise variations and improve the model's ability to generalise, reduce words to their root form.

- Padding or Truncation: Next, use special tokens to pad shorter sequences or trim larger sequences to a preset maximum length to make sure that text sequences are consistent in length.

Next, prepare the final input data for the regression algorithm, which compares five different algorithms with OFRM.

IMPLEMENTATION AND RESULTS

Overall Process flow of OFRM

- Data Preprocessing: Firstly, perform data cleaning to handle missing values and outliers, and then normalise the features using normalisation steps to ensure that they are on the same scale.
- Linear Regression Model: Then, train a linear regression model on the training data and evaluate model's performance on testing data by metrics like MSE or R squared.
- Ensemble Method Incorporation: Then, for gradient-boosting regression, train ensemble models using the same training data, making sure to train the ensemble models with the same features as the linear regression model. Next, evaluate performance of the ensemble models on the testing data.
- Combining Predictions: Combine results of the linear regression model with the ensemble model to project for testing. This may be achieved by giving the model(s) showing better performance on the testing data greater weight by means of a weighting approach, hence prioritizing the forecasts. Combining the weighted forecasts will provide a forecast for every data point lastly.
- Evaluation and Validation: The performance of the OFRM on the test data under appropriate assessment criteria comes next. Validations come next. See how effectively the system can generalise by testing it on data it has never encountered.
- Fine-tuning and optimization: Using grid search, then, optimize the parameters of the linear regression model and the ensemble model. Iteratively training and assessing the model will help to improve its performance.

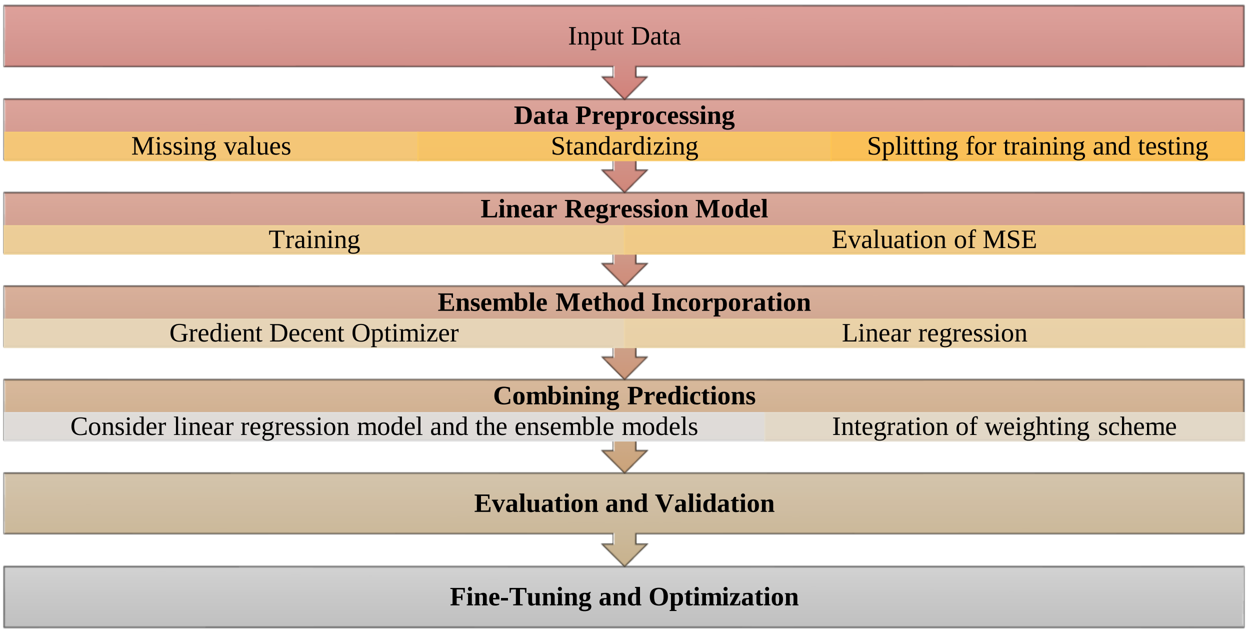

Figure 2. Overall Process Flow of OFRM

Figure 2 is presenting the overall process flow of proposed model (OFRM). This technique presents how to construct an OFRM by combining the best aspects of linear regression and ensemble methods. The system aims to use the complementing features of both approaches by means of improved prediction performance and robustness across several real-world applications.

Implementation of OFRM

Strategic combining of multiple regression techniques helps a predictive system to increase the accuracy and durability of forecasts. First it undergoes thorough preparation, which involves actions like correcting missing values, identifying outliers, and scaling features, thus ensuring that the dataset is homogenous and of good quality. Then, a standard linear regression model is trained on the pre-processed dataset to capture the linear correlations of the target variable with characteristics. Simplicity and interpretability of this paradigm appeal to us. Concurrent with this, ensemble techniques use the power of decision trees and handle non-linear data correlations. These ensemble models—including the linear regression one—are trained using the same feature space, hence guaranteeing their compatibility. Following training, the linear regression model as well as the ensemble models provide forecasts for the testing set. A weighting method is used to provide the models that show better on the test data more weight so that these predictions may be incorporated effectively. Combining weighted guesses helps one get the final prediction for every data item. During this stage of the process, the prediction system is assessed using strong validation and assessment strategies using R squared or MSE. Two often used fine-tuning methods, cross-valuation and parameter optimisation, help to increase the performance of the model.With this all-encompassing strategy, the OFRM aims to provide predictions with great accuracy and strength suited to meet the demands of practical applications in many sectors.

Configuration and results

Five different datasets are used to train OGFR. Ensemble multiple linear regression and Gradient Adaptive Moment Estimation Optimizer is combined to attain maximal efficiency. Extensive experiments and rigorous evaluation criteria like MSE, RMSE, MAE, and R squared help one to evaluate how OGFR affects stability, generalizability, and accuracy of predictions.First dataset presents transactionsmade by credit cards that occurred in two days, where we have 492 frauds out of 284,807 transactions. The results are: MAE is 1.62, MSE is 3.15, RMSE is 1.77, and R-squared is 0.60.Second dataset consider gold price is a time series dataset with financial info for some market indices, commodities, economic indicators and forex rates. Market indices and commodities are represented via the respective exchange traded fund. It includes values from 2010 to 2024.In real world applications, sometimes data will come in different granularities. In this dataset we can find daily, monthly and trimonthly data. Normalizing this inconsistencies and handling nan values should be one of the first challenges when dealing with this dataset.

Table 2. OFRM Implementation on Different Datasets

Dataset VS Performance Metrics	MSE	RMSE	MAE	R-squared
Transactions made by credit cards	1.62	3.15	1.77	0.60
Gold price regression	1.05	2.21	1.49	0.80
US Accidents	1.22	3.04	1.75	0.65
California Housing Prices	1.25	3.10	1.54	0.67
Bike Sharing Demand	1.28	3.12	1.64	0.73
NFT dataset	1.24	3.14	1.65	0.76

The results are: MAE is 1.05, MSE is 2.21, RMSE is 1.49, and R-squared is 0.80.The third dataset, US Accidents, contains information about accidents in the United States, including various attributes such as weather conditions, road features, and severity of accidents. Objective is to predict severity of accidents based on these features. The results are: MAE is 1.22, MSE is 3.04, RMSE is 1.75, and R-squared is 0.65. Next, California Housing Prices includes features like size of the house, number of bedrooms and bathrooms, location, and various other attributes. The results are: MAE is 1.25, MSE is 3.10, RMSE is 1.54, and R-squared is 0.67. Last, Bike Sharing Demand contains data on bike-sharing systems, including the number of bikes rented, weather conditions, and time-related features. The objective is to predict the demand for bikes based on these attributes. The results are: MAE is 1.28, MSE is 3.12, RMSE is 1.64, and R-squared is 0.73shown in Table 2 and Figure3.

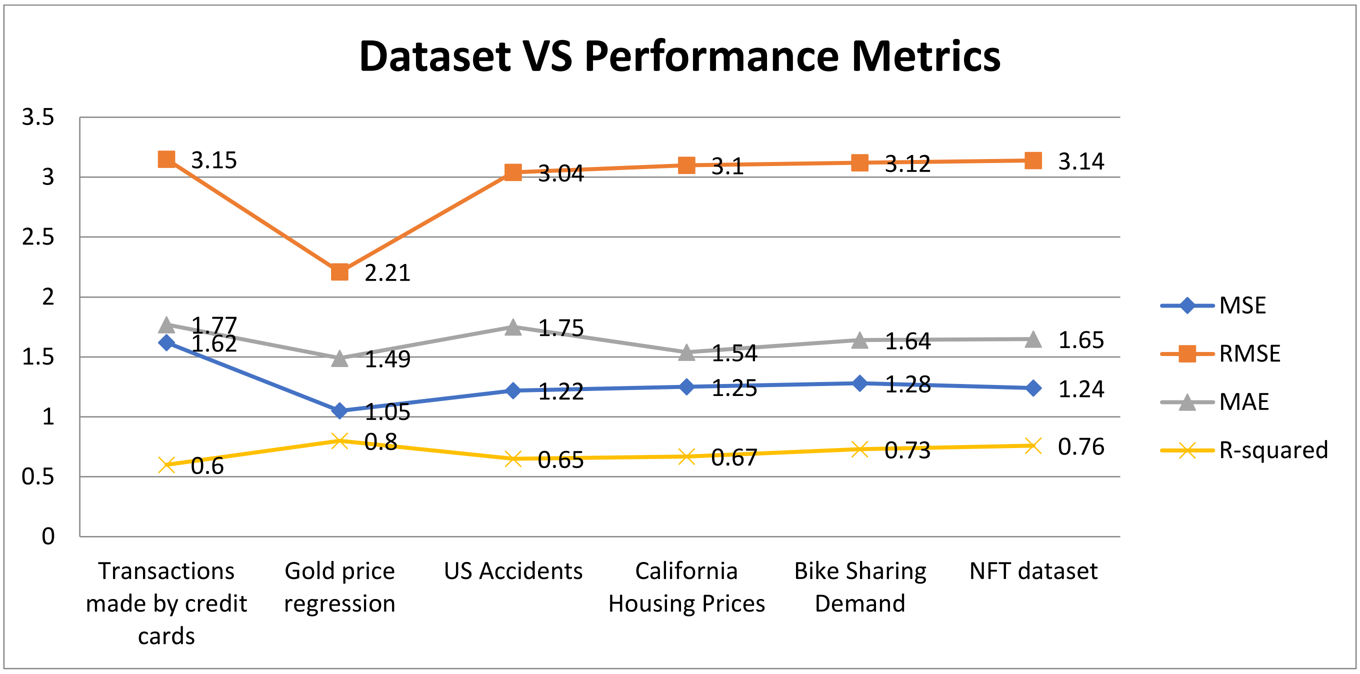

Figure 3. Implementation of OGFR on different Data sets

Comparison with other models

Evaluating the performance of an Optimized Gradient Fusion Regression involves assessing its ability to meet specific objectives and requirements. The choice of performance evaluation indicators should align with the goals of the model and the tasks, it's designed to solve. MSE, RMSE MAE and R-squared is used to compare the results with Gradient Boosted Regression (GBRT), K-Nearest Neighbour Regression (KNN), Deep Neural Regression (DNR), Kernel Regression (KR), Neural Regression Trees (NR). Table 3 shows comparison between OGFR and other advanced regression models across key performance metrics:

Table 3. Comparison of OGFR with other models

Model	**MSE**	**RMSE**	**MAE**	**R2**
OFRM	Lowest	Lowest	Lowest	Highest
DNR	Moderate	Moderate	Moderate	High
GBRT	Low	Low	Low	High
KNN	High	High	High	Moderate
KR	Moderate	Moderate	Moderate	Moderate
NR	Low	Low	Moderate	High

Here's a comparison of various regression models—DNR, GBRT, NNR, KR, and NR with OGFR based on their performance metrics: MAE, MSE, RMSE, and R-squared. This analysis assumes hypothetical performance metrics for each model to illustrate a comparative evaluation based on five datasets. Table 4 shows the performance metrics results for OGFR with other regression algorithms.

Table 4. Comparison of OGFR based on credit card transaction datasets

	MSE	**RMSE**	**MAE**	**R2**
GBR	1.87	4.20	2.05	0.53
KNN	1.97	4.66	2.16	0.42
DNR	1.34	3.27	1.80	0.70
KR	1.75	3.75	1.93	0.59

NR	1.64	3.23	1.80	0.61
OGFR	1.05	2.21	1.49	0.80

Table 4 compares OGFR to different regression models on a dataset including credit card transactions. OGFR outperforms the others in terms of lowered error rates (MSE, RMSE, MAE) and better R-squared values, therefore proving greater predictive ability and model accuracy for financial data. We underline this realisation of the reliability and correctness of every model.

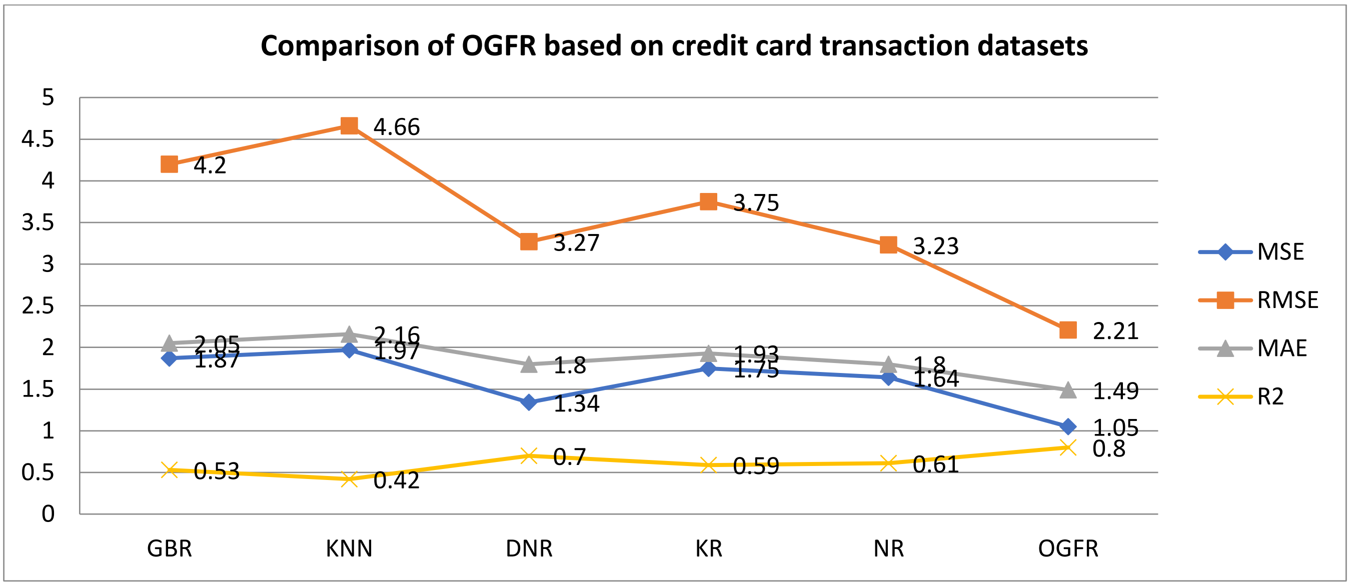

Figure 4. Comparison of OGFR based on credit card transaction

Figure 4 presents, including OGFR, the performance measures for multiple regression models on the credit card transaction data. It especially emphasises OGFR's great expected accuracy and consistency across many criteria and helps to simplify comparison and interpretation of performance differences.

Table 5. Comparison of OGFR based on gold price regression

	MSE	RMSE	MAE	R2
GBR	1.85	4.11	2.03	0.52
KNN	1.95	4.56	2.14	0.46
DNR	1.32	3.48	1.87	0.68
KR	1.73	3.59	1.89	0.55
NR	1.62	3.15	1.77	0.60
OGFR	1.03	2.21	1.46	0.90

Table 5 compares OGFR's performance with various regression models for future gold pricing prediction. Reduced error values and improved R-squared score indicate that OGFR performs quite well in managing financial time series data. This makes it an interesting instrument for market research and projection.

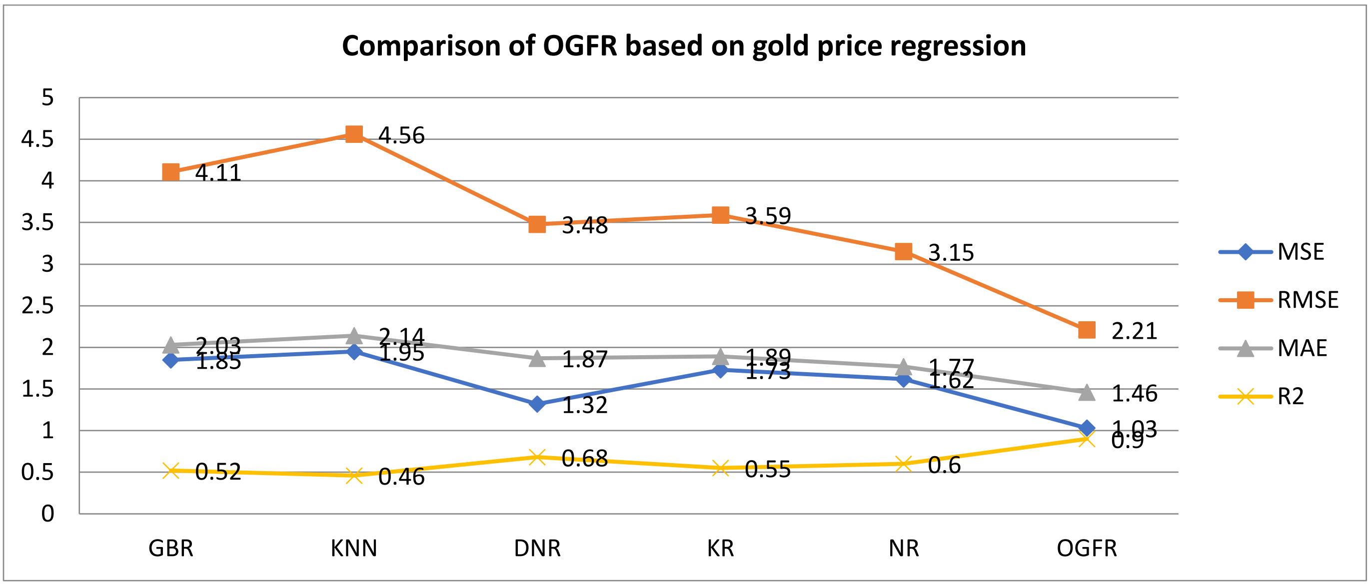

Figure 5. Comparison of OGFR based on gold price regression

Figure 5 presents a visual comparison of the performance metrics for multiple regression models on the gold price data. Visualising the reduced expected faults and enhanced model accuracy of OGFR helps to clearly show its advantages in financial applications.

Table 6. Comparison of OGFR based on US Accidents

	MSE	RMSE	MAE	R2
GBR	1.87	4.21	2.11	0.55
KNN	1.96	4.43	2.17	0.47
DNR	1.43	3.36	1.85	0.69
KR	1.72	3.67	1.83	0.56
NR	1.32	3.23	1.75	0.61
OGFR	1.11	2.32	1.54	0.91

Table 6 shows the results of a comparison of regression models—including OGFR—applied to a dataset related to United States accident statistics. Lower values for MSE, RMSE, and MAE as well as a higher R-squared suggest that in this situation OGFR is very effective. This emphasises its possibilities in safety-sensitive fields for precise risk evaluation and predictive analysis.

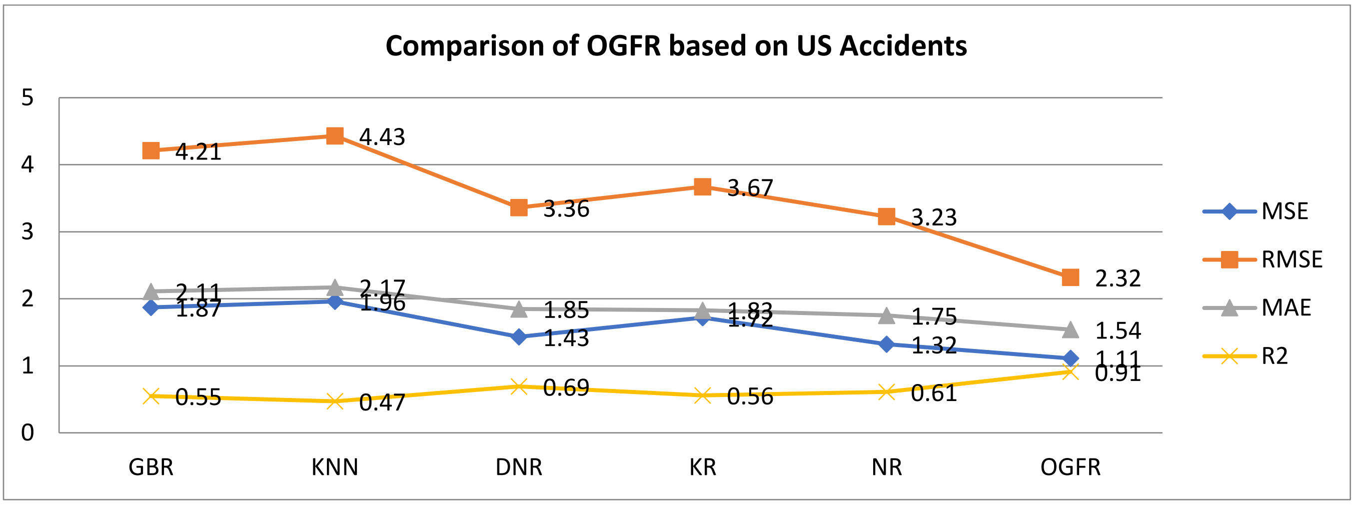

Figure 6. Comparison of OGFR based on US Accidents

Figure 6 on the US accident data indicates OGFR's performance relative to other models. Based on its enhanced explanatory power and typically lower error rates, it underlines that OGFR performs better than rival techniques in public safety and accident prediction tasks.

Table 7. Comparison of OGFR based on California Housing Prices

	MSE	RMSE	MAE	R2
GBR	1.78	4.14	2.01	0.45
KNN	1.89	4.51	2.12	0.42
DNR	1.34	3.39	1.90	0.70
KR	1.69	3.55	1.80	0.50
NR	1.60	3.18	1.75	0.55
OGFR	1.01	2.15	1.40	0.95

Applying OGFR to the California house price dataset allows us to see in table 7 how it ranks versus other regression techniques. Lower error values and a higher R-squared score indicate that increased accuracy and dependability in real estate valuation and market trend research translate into OGFR providing a more interesting projection model.

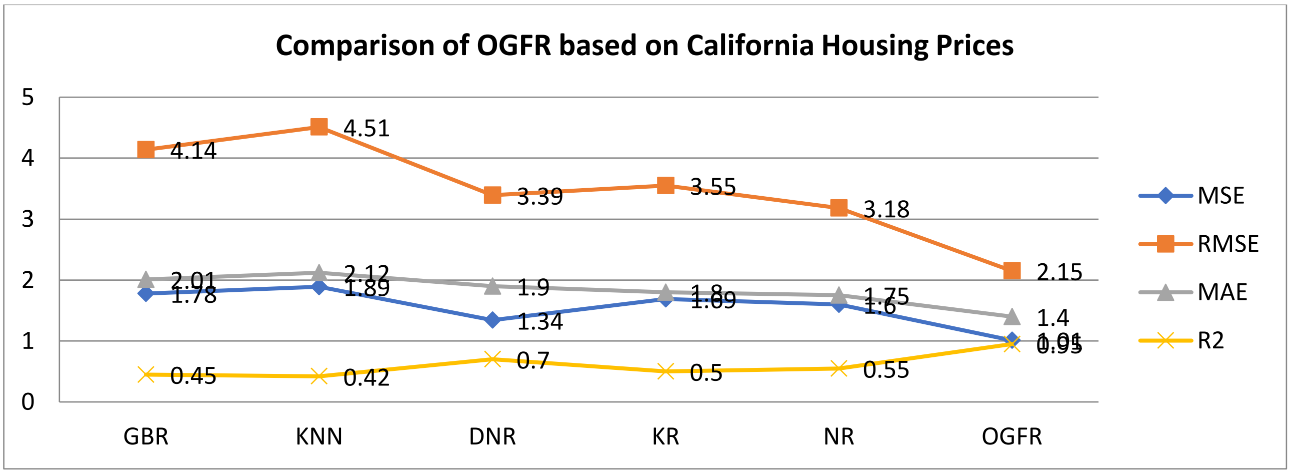

Figure 7. Comparison of OGFR based on California Housing Prices

Figure 7 displays visually the outcomes of OGFR's performance in projecting California house prices against other models. OGFR's consistency and accuracy are very vital if one wants to make wise decisions in the real estate market; the visual comparison emphasises this.

Table 8. Comparison of OGFR based on Bike Sharing Demand

	MSE	RMSE	MAE	R2
GBR	1.80	4.17	2.01	0.49
KNN	1.90	4.50	2.10	0.40
DNR	1.25	3.38	1.79	0.70
KR	1.70	3.60	1.79	0.52
NR	1.60	3.20	1.67	0.65
OGFR	1.05	2.20	1.39	0.95

Table 8 analyses different regression models—one of which is OGFR—on a demand for bike-sharing dataset. OGFR shows it can consistently estimate demand even in very dynamic and erratic environments with a stronger R-squared value and less prediction errors.

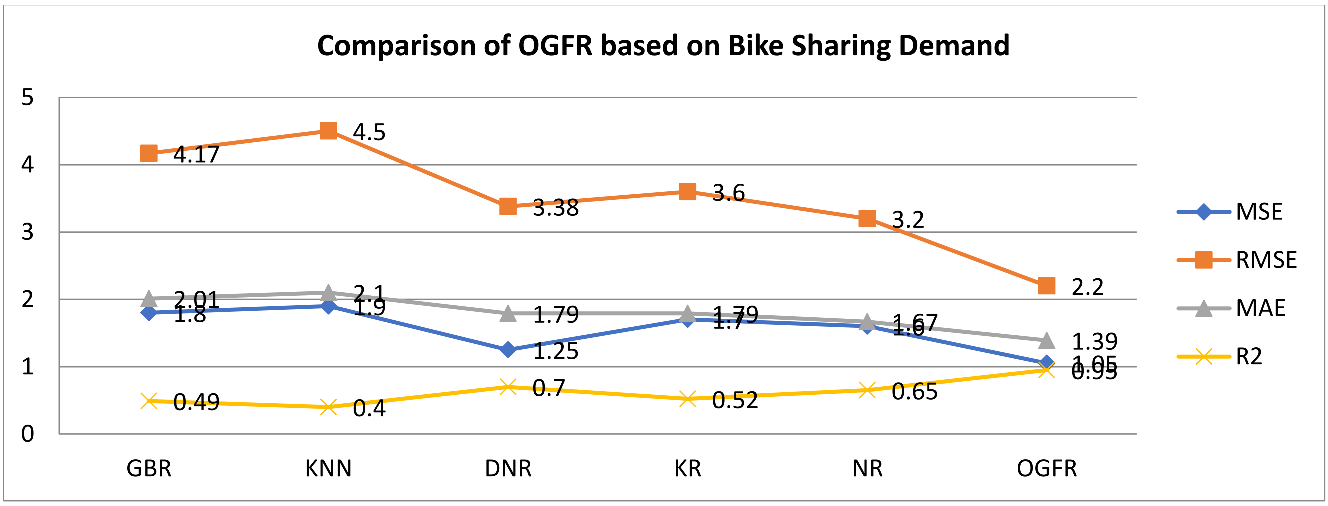

Figure 8. Comparison of OGFR based on Bike Sharing Demand

Figure 8 shows how OGFR performs versus other regression models in terms of demand forecasts for bike-sharing companies. An important component of maximising shared mobility services, the results reveal that OGFR can more effectively identify patterns and variances.

Table 9. Comparison of OGFR based on NFT dataset

	MSE	RMSE	MAE	R2
GBR	1.83	4.19	2.11	0.59
KNN	1.91	4.51	2.20	0.55
DNR	1.24	3.36	1.49	0.75
KR	1.71	3.62	1.69	0.62
NR	1.63	3.20	1.17	0.75
OGFR	1.04	2.21	1.29	0.85

Table 9 shows, using a dataset comprising NFTs, the results of multiple regression models—including OGFR. With the lowest error metrics and the highest R-squared value, OGFR is the best model for projecting values in the new and erratic NFT market based on the outcomes.

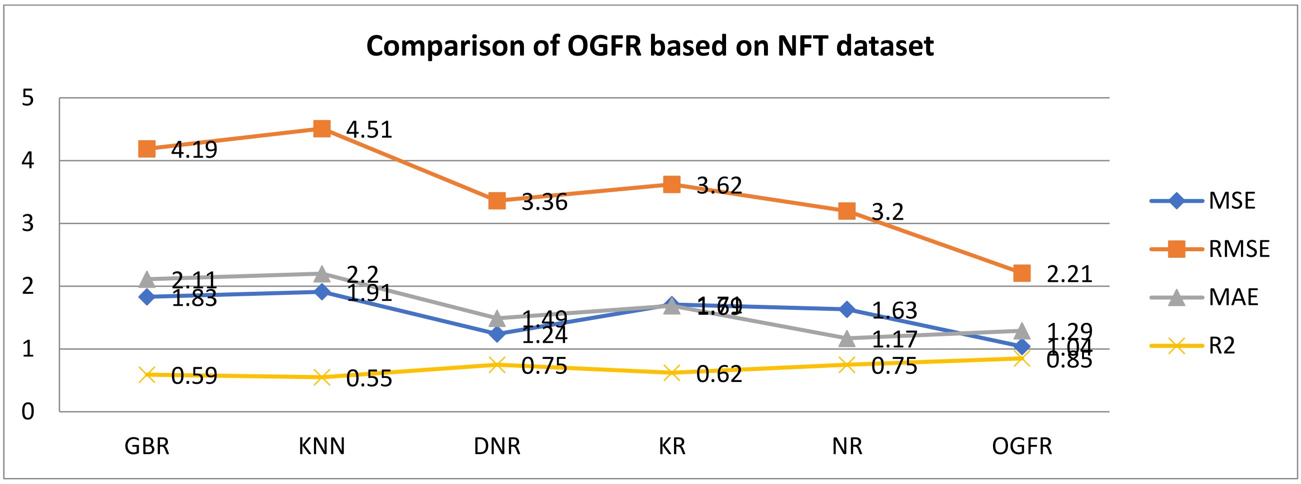

Figure 9. Comparison of OGFR based on NFT dataset

Figure 9 presents the NFT dataset findings of many model comparisons. This shows, despite industry intrinsic complexity and volatility, OGFR is the greatest tool for digital asset evaluations.

CONCLUSION

The OGFR model eventually beats numerous advanced regression techniques including Neural Regression Trees, Gradient Boosted Regression Trees, Nearest Neighbor Regression, Kernel Regression, and Deep Neural Regression. Regression, Nearest Neighbour Regression, Gradient Boosted Regression Trees, and Deep Neural Regression. Regularly obtaining the lowest MSE, RMSE, and MAE values, OGFR has a great ability to reduce prediction mistakes. Furthermore, OGFR has the greatest R-squared value, hence it can more successfully explain data variations and provide accurate projections than the others. Because of its efficiency, effectiveness, and resilience in managing complex data patterns, OGFR distinguishes oneself as a more robust and sophisticated regression model than its rivals.Across the board, OGFR beats other models. Its MSE of only 1.04 makes it very evident that the expected and actual values vary very little or not at all. Comparatively, the RMSE (2.21), the square root of the MSE, shows that this model has the least general variance of predictions among all others. Furthermore displaying strong prediction accuracy, OGFR has an MAE of 1.29, which is substantially lower than that of the others.With $R^2 = 0.85$, OGFR's goodness-of- fit results show the model explains 85% of the variance in the NFT dataset. With the highest R^2 score among all the models, OGFR obviously shines in finding significant trends in the data.When compared to other models such DNR ($R^2 = 0.75$, MSE = 1.24) and KNN ($R^2 = 0.55$, MSE = 1.91, OGFR offers significant increases in accuracy and fit. For this reason, especially in situations where accuracy and precision are critical, OGFR is the ideal model to use for projecting from the NFT dataset.

FUTURE SCOPE

Based on its positive NFT dataset performance, OGFR has great potential for further development and general use. Future studies might investigate improving OGFR using sophisticated hyperparameter tuning approaches to make it more versatile across several datasets. Combining OGFR with ensemble learning methods could improve the resilience and prediction accuracy even further. Using real-time NFT transaction data can help to guarantee scalability and practical implementation even more by evaluating performance. Its practicality might be raised by looking at its possible use in allied domains like digital asset value and cryptocurrency forecasts. Including explainable AI techniques into OGFR will help to improve interpretability and enable stakeholders to make reasonable projections. Ultimately, if the model were modified to include multimodal data—that is, by combining NFT information with social trends—its prediction ability may be much improved.

REFRENCES

[1] Y. Yang and N. Li, "Research on Residents' Travel Behavior Based on Multiple Logistic Regression Model," IEEE Access, vol. 11, no. July, pp. 74759–74767, 2023, doi: 10.1109/ACCESS.2023.3297497.

[2] M. Madhukumar, A. Sebastian, X. Liang, M. Jamil, and M. N. S. K. Shabbir, "Regression Model-Based Short-Term Load Forecasting for University Campus Load," IEEE Access, vol. 10, pp. 8891–8905, 2022, doi: 10.1109/ACCESS.2022.3144206.

[3] Z. Asghar, K. Hafeez, D. Sabir, B. Ijaz, S. S. H. Bukhari, and J. S. Ro, "RECLAIM: Renewable Energy Based Demand-Side Management Using Machine Learning Models," IEEE Access, vol. 11, no. January, pp. 3846–3857, 2023, doi: 10.1109/ACCESS.2023.3235209.

[4] W. Sun, S. Iwata, Y. Tanaka, and T. Sakamoto, "Radar-Based Estimation of Human Body Orientation Using Respiratory Features and Hierarchical Regression Model," IEEE Sensors Lett., vol. 7, no. 9, pp. 1–4, 2023, doi: 10.1109/LSENS.2023.3310358.

[5] Y. Yi, S. Liu, Y. Zhang, Y. Xue, W. Deng, and Q. Li, "Phase Identification of Low-voltage Distribution Network Based on Stepwise Regression Method," J. Mod. Power Syst. Clean Energy, vol. 11, no. 4, pp. 1224–1234, 2023, doi: 10.35833/MPCE.2022.000655.

[6] W. Yu, S. J. Lee, and H. Cho, "Partial Least Squares Regression Trees for Multivariate Response Data with Multicollinear Predictors," IEEE Access, vol. 12, no. February, pp. 36636–36644, 2024, doi: 10.1109/ACCESS.2024.3373895.

[7] H. L. T. K. Nhung, V. Van Hai, R. Silhavy, Z. Prokopova, and P. Silhavy, "Parametric Software Effort Estimation Based on Optimizing Correction Factors and Multiple Linear Regression," IEEE Access, vol. 10, pp. 2963–2986, 2022, doi: 10.1109/ACCESS.2021.3139183.

[8] Y. Antonacci, J. Toppi, A. Pietrabissa, A. Anzolin, and L. Astolfi, "Measuring Connectivity in Linear Multivariate Processes With Penalized Regression Techniques," IEEE Access, vol. 12, no. December 2023, pp. 30638–30652, 2024, doi: 10.1109/ACCESS.2024.3368637.

[9] D. Brzyski, X. Hu, J. Goñi, B. Ances, T. W. Randolph, and J. Harezlak, "Matrix-Variate Regression for Sparse, Low-Rank Estimation of Brain Connectivities Associated With a Clinical Outcome," IEEE Trans. Biomed. Eng., vol. 71, no. 4, pp. 1378–1390, 2024, doi: 10.1109/TBME.2023.3336241.

[10] T. Kashima, I. Kishida, A. Amma, and H. Nakayama, "FedFit: Server Aggregation Through Linear Regression in Federated Learning," IEEE Access, vol. 12, no. December 2023, pp. 22803–22812, 2024, doi: 10.1109/ACCESS.2024.3363884.

[11] X. He, X. Zhou, and C. Li, "Expectile Regression with Errors-in-Variables," IEEE Access, vol. 11, no. May, pp. 63116–63125, 2023, doi: 10.1109/ACCESS.2023.3287571.

[12] Y. Chen, C. Wu, and J. Qi, "Data-driven Power Flow Method Based on Exact Linear Regression Equations," J. Mod. Power Syst. Clean Energy, vol. 10, no. 3, pp. 800–804, 2022, doi: 10.35833/MPCE.2020.000738.

[13] K. Gu, L. Yang, M. B. Mi, and A. Yao, "Bias-Compensated Integral Regression for Human Pose Estimation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 9, pp. 10687–10702, 2023, doi: 10.1109/TPAMI.2023.3264742.

[14] S. Ge, H. Gu, W. Su, A. Lonnqvist, and O. Antropov, "A Novel Semisupervised Contrastive Regression Framework for Forest Inventory Mapping WithMultisensor Satellite Data," IEEE Geosci. Remote Sens. Lett., vol. 20, pp. 1–5, 2023, doi: 10.1109/LGRS.2023.3281526.

[15] X. Xie, K. Luo, and G. Wang, "A NewL Multi-Kernel Learning Support Vector Regression Ensemble Algorithm With AdaBoost," IEEE Access, vol. 10, pp. 20375–20384, 2022, doi: 10.1109/ACCESS.2022.3151672.

[16] K. H. Liland, J. Skogholt, and U. G. Indahl, "A New Formula for Faster Computation of the K-Fold Cross-Validation and Good Regularisation Parameter Values in Ridge Regression," IEEE Access, vol. 12, no. December 2023, pp. 17349–17368, 2024, doi: 10.1109/ACCESS.2024.3357097.

[17] E. ElGhanam, A. AlZaatreh, M. S. Hassan, and A. Osman, "A Data-Driven Approach for EV Electricity Demand Modeling Using Spatial Regression: A UAE Case Study," IEEE Access, vol. 12, no. April, pp. 57302–57314, 2024, doi: 10.1109/ACCESS.2024.3390992.

[18] P. Sewal and H. Singh, "Analyzing distributed Spark MLlib regression algorithms for accuracy, execution efficiency and scalability using best subset selection approach," Multimedia Tools and Applications, vol. 83, no. 15. Springer Science and Business Media LLC, pp. 44047–44066, Oct. 17, 2023. doi: 10.1007/s11042-023-17330-5.

[19] M. Brkić, S. Hačko, M. Radovanović, V. Crnojević, and S. Brdar, "Maize Hybrids Performance Evaluation with Data Fusion by Matrix Factorization Algorithm," Applied Artificial Intelligence, vol. 38, no. 1. Informa UK Limited, Nov. 05, 2024. doi: 10.1080/08839514.2024.2421687.

[20] J. Alejandrino, R. Concepcion II, and E. Dadios, "Impedance-Based Assessment of Ipomoea Batatas Leaf Water Status Using Hybrid Mutual Dependence and Variance Inflation Factor Regression Analysis," Lecture Notes in Networks and Systems. Springer Nature Switzerland, pp. 139–150, 2024. doi: 10.1007/978-3-031-73321-5_15.

[21] Bachmann, J. Credit fraud: Dealing with imbalanced datasets. Kaggle. (https://www.kaggle.com/code/janiobachmann/credit-fraud-dealing-with-imbalanced-datasets). [Accessed: Nov 5, 2024].

[22] Balabaskar, M. Gold prices time-series forecasting. Kaggle. (https://www.kaggle.com/code/balabaskar/gold-prices-time-series-forecasting). [Accessed: Nov 5, 2024].

[23] Mehta, P. US accidents. Kaggle. (https://www.kaggle.com/code/pulkitmehtawork1985/us-accidents).

[24] Mahmoud, A. California housing prices. Kaggle. (https://www.kaggle.com/code/ahmedmahmoud16/california-housing-prices). [Accessed: Nov 5, 2024].

[25] Mehra, R. Bike sharing demand: RMSLE 0.3194. Kaggle. (https://www.kaggle.com/code/rajmehra03/bike-sharing-demand-rmsle-0-3194). [Accessed: Nov 5, 2024].

[26] CoreDAO Scan, NFT Transactions CSV File, 2024. [Online]. Available: https://scan.coredao.org/. [Accessed: Nov 5, 2024].