Explainable and Fair AI: Balancing Performance in Financial and Real Estate Machine Learning Models

Article *in* IEEE Access · November 2024

DOI: 10.1109/ACCESS.2024.3484409

3 authors:

Deepak Acharya
University of Alabama in Huntsville
12 PUBLICATIONS **294** CITATIONS
SEE PROFILE

Divya B Ashwin
Manipal Academy of Higher Education
7 PUBLICATIONS **191** CITATIONS
SEE PROFILE

Karthigeyan Kuppan
JPMorgan Chase
9 PUBLICATIONS **182** CITATIONS
SEE PROFILE

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 October 2024, accepted 17 October 2024, date of publication 22 October 2024, date of current version 30 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3484409

RESEARCH ARTICLE

Explainable and Fair AI: Balancing Performance in Financial and Real Estate Machine Learning Models

DEEPAK BHASKAR ACHARYA[1], (Senior Member, IEEE), B. DIVYA[2], AND KARTHIGEYAN KUPPAN[3], (Senior Member, IEEE)

[1]Department of Computer Science, The University of Alabama in Huntsville, Huntsville, AL 35806, USA
[2]Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India
[3]JPMorgan Chase & Company, Houston, TX 77002, USA

Corresponding author: B. Divya (divya.ashwin@manipal.edu)

ABSTRACT This paper introduces a framework that integrates fairness and transparency into advanced machine learning models, specifically LightGBM and XGBoost, applied to loan approval and house price prediction datasets. The key contribution is using fairness-focused techniques, such as Calibrated Equalized Odds and Intersectional Fairness, which are not widely studied in financial and real estate contexts. To improve model transparency, SHAP (SHapley Additive exPlanations) is utilized along with a novel fairness-based interpretability method to measure both model fairness and the importance of individual features. Through comprehensive experiments, we show that LightGBM delivers high accuracy while balancing fairness and performance effectively. The broader relevance of this work is discussed in the context of governance and regulatory requirements, highlighting the importance of responsible practices in high-stakes financial decision-making processes. This research highlights the importance of fairness and transparency in real-world applications, promoting equity, trust, and adherence to evolving legal standards, and provides practical insights for data scientists, machine learning researchers, and professionals in the real estate and financial sectors.

INDEX TERMS Fairness in AI, explainable AI, SHAP, loan approval prediction, equalized odds, intersectional fairness, LightGBM, XGBoost, AI governance.

I. INTRODUCTION

Machine learning models are at the heart of the finance and real estate [1] decision-making processes. The use of these models can increase accuracy; however, issues tied to fairness and transparency remain. Many models can be biased, resulting in unethical decisions that negatively affect some sections of society. Also, explaining these models is imperative in building confidence in their applications, especially in industries subject to enforcement.

While the concepts of fairness [2] and explainability have been explored quite a bit in healthcare and criminal justice, they are not so much accepted and practiced in the case of financial and real estate decision-making. These industries undertake weighty decision-making that shapes people's access to ad resources; for instance, U.S. citizens have biased models regarding loan approvals and housing prices, and the consequences of concentrated disadvantage can be alarming. Therefore, it is apparent that such issues have to be resolved. More importantly, AI systems adopted in the finance and real estate industries will not have the biases contained in the training datasets. Also, regulatory pressure has increased; with it, there is an increasing need for the models to be right, understandable, and fair.

Recent developments regarding explainable artificial intelligence (XAI) methodologies, like SHAP, PDPs and ALE, have been shown to enhance model transparency and allow for fairness in decision making across different fields including finance, medicine, and real estate. These types of approaches have also been fruitful in the field

The associate editor coordinating the review of this manuscript and approving it for publication was Jolanta Mizera-Pietraszko.

of medicine, especially in areas like image segmentation, for example polyp detection in colonoscopy images [3] where explanation and precision are crucial for the diagnosis and treatment. Further advancements have extended these techniques to community detection and clustering tasks using Gumbel-Softmax in graph neural networks, showing significant improvements in performance across various network datasets [4], [5], [6]. Similarly, ensemble learning models have proven to be effective in complex medical imaging tasks, such as brain tumor segmentation [7], [8], highlighting the growing importance of explainable models in high-stakes environments.

By explaining what features contributed most to model performance, XAI [9] has become timely in solving the need for more transparency in model predictions. More recent work has demonstrated the use of XAI methods, such as SHAP, PDPs, and ALE, in modeling rent predictions and other financial models [10], [11]. These techniques assist other relevant parties in viewing which features of a person, e.g., income, credit history, type of property, have a bearing on the outcome, and so such discrepancies or inadvertent inequities can be spotted.

This research tries to fill a major gap by targeting the fairness constraints and explainability tools in financial and real estate systems, which remain less explored than the other fields in the literature. There is a great need to avoid the accuracy and fairness trade-off in these high-stakes sectors, as this would help to advance equitable opportunities and resources. By using fairness-aware machine learning techniques, this research seeks to offer adequate interventions that can easily be implemented in practice in a manner that adheres to regulations and also supports ethical AI systems.

In this paper, we provide a detailed analysis of two widely used gradient boosting models of classification and regression analysis, LightGBM and XGBoost [12], for declining or approving a loan request and forecasting housing prices. We assess the accuracy versus fairness versus interpretability dilemma through fairness integrating methods that include Calibrated Equalized Odds, Intersectional Fairness, and transparency, such as SHAP.

II. LITERATURE REVIEW

A. FAIRNESS IN AI

Fairness in machine learning has been a growing concern, mainly when models are deployed in high-stakes finance, healthcare, and law enforcement. Recent work by [13] highlights the ongoing developments at the frontiers of fairness in machine learning, providing insights into challenges and emerging solutions. Discriminatory outcomes can arise when models inadvertently reflect biases present in the training data, leading to unequal treatment of certain demographic groups. Reference [14] extensively reviewed existing and proposed bias mitigation strategies in the machine learning domain, articulating ways to achieve fairness [15] in various contexts. In particular, [16] assessed the effectiveness of fairness-enhancing algorithms on credit scoring and loan approval decisions to reduce biases. Reference [17] offered a synthesis of bias-aware algorithms, particularly in machine learning, and the practical issues that need to be addressed when these algorithms are used in high television applications.

In addition, [11] and [18] has provided a comprehensive review of XAI techniques in the financial domain, illustrating how bias in credit scoring and fraud detection models can be mitigated through fairness-aware learning frameworks. This demonstrates the increasing importance of ethical considerations in finance [19], especially when decisions are made using AI models that have significant real-world consequences. Further research by [20] investigated the role of fairness-aware deep learning models, particularly in fraud detection and algorithmic credit scoring.

B. EXPLAINABILITY IN AI

Explainability has become a crucial component of AI models, especially in domains where trust is vital. As explored in [21], interpretability in machine learning has evolved over the years, with recent efforts focusing on balancing accuracy with transparency. Black-box models, such as deep neural networks and gradient-boosting machines, often exhibit high performance but lack interpretability. Reference [22] discussed the rising importance of explainable AI (XAI) techniques, particularly in sensitive applications where understanding model behavior is essential. In addition, [23] provided insights into using XAI to enhance the transparency of AI models in financial risk assessment, allowing decision-makers to understand the rationale behind model predictions.

Furthermore, [10] analyzed multiple XAI methods in real estate applications, particularly SHAP and ALE, to provide local and global explanations for rent predictions. These techniques can be adapted to financial models to ensure transparency in decision-making, as they offer insights into how non-linear relationships between features like income or property size affect predictions. Reference [24] further highlighted how XAI methods can be used in real estate to ensure transparent decision-making, particularly in property investment scenarios. Reference [25] explored counterfactual explanations, adding another dimension to explainability by showing how changes to specific inputs would have led to different predictions, enhancing trust in AI models.

C. MACHINE LEARNING MODELS IN FINANCE AND REAL ESTATE

Machine learning has been increasingly applied to financial and real estate sectors, where accurate predictions can significantly impact economic outcomes. In finance, machine learning models are used for credit scoring, fraud detection, and loan approval decisions. Traditional credit scoring models have relied on linear regression and decision trees, but more recent approaches incorporate gradient boosting machines and neural networks [26]. Reference [27] discussed the efficiency of using gradient-boosting machines for credit scoring, which is better than the conventional approach. Also, [28] looked

into the deployment of neural networks in various domains, including the prediction of loan defaults, where better results than traditional models were obtained. In [29], the authors also investigated the applicability of several ensemble methods for credit risk prediction using XGBoosted, LightGBM, and Random Forest models to increase prediction accuracy with less bias to an individual model.

In real estate, machine learning is used to predict house prices [30], evaluate property investments, and assess market trends. Gradient boosting algorithms like XGBoost and LightGBM are particularly well-suited for real estate predictions because they handle non-linear relationships and high-dimensional data. Reference [31] demonstrated the effectiveness of XGBoost in predicting housing prices, outperforming traditional methods such as linear regression. The authors of [32] created a real estate price prediction model that combines neural networks with XGBoost [33], focusing on nonlinearity and feature interactions. Additionally, [34] investigated the use of machine learning in predicting rental prices, applying some XAI features as well when synergizing location, size, and amenities.

The systematic review by [11] has highlighted the role of explainability in financial applications, with SHAP, LIME, and other XAI methods gaining traction in loan approval systems and mortgage lending. These XAI techniques provide stakeholders with the necessary transparency to comply with regulatory requirements, particularly in mitigating biases related to demographic attributes such as gender and race. Reference [35] further explored the role of XAI in improving transparency in financial fraud detection, highlighting how it can be used to explain complex fraud patterns to financial auditors and regulators.

III. METHODOLOGY

This research follows a more detailed framework design incorporating a fairness approach and explainability methods into the machine learning models for financial and real estate decisions. The two major tasks include predicting the probability of sanctioning the loan and predicting the amount of house going to be sold in the market. The methodology incorporates data preprocessing, model training, fairness constraints and explanability techniques.

A. DATA COLLECTION AND PREPROCESSING

The datasets used in this study are sourced from publicly available financial and real estate datasets:

- **Loan Approval Dataset**: This dataset comprises details such as applicant's income, loan amount, credit history and other demographic aspects.
- **House Price Prediction Dataset**: Some of the features contained in this dataset include overall quality, living area, neighborhood, property and any other relevant variables.

Both of the datasets were subjected to general preprocessing steps including:

- **Data Cleaning**: This information was processed, such that missing values, where applicable, were handled by mean in case of numerical features whereas mode for categorical features was preferred.
- **Feature Engineering**: Additional features, such as loan-to-income ratio and property location features, were created to enhance model predictions.
- **Data Normalization**: These numerical features were normalized so as to enhance the performance of gradient-boosting models.

B. MODEL SELECTION

For both tasks, two powerful gradient boosting algorithms were used:

- **LightGBM**: Known for its efficiency in handling large datasets and fast training times.
- **XGBoost**: A widely used gradient boosting model that is effective for structured data.

Both models were trained using 80% of the data, with 20% set aside for testing. Hyperparameter tuning was performed using grid search to optimize parameters such as learning rate, number of trees, and max depth.

C. FAIRNESS CONSTRAINTS

Fairness constraints were built into the models in order to promote fairness in the decision-making process for the different demographic groups considered in the models, such as the gender and income level for loan approvals or location and property size for house pricing. The following fairness metrics were used:

- **Calibrated Equalized Odds**: Assures the equal discriminatory rates for false positive and false negative across different sections.
- **Intersectional Fairness**: Fairness concerning the intersection of more than one protected attribute, such as gender and income, as well as education and race is analyzed

These fairness constraints were added to the models using post-processing methods, where model outputs are modified to fulfill the fairness criterion without much loss of accuracy.

D. EXPLAINABILITY WITH SHAP AND XAI TOOLS

To ensure that our models provided both accurate predictions and meaningful explanations, SHAP (SHapley Additive exPlanations) was employed to evaluate the contribution of individual features to the model outputs. SHAP values were computed for all features in both datasets, providing insights into how each feature contributed to the final prediction. In line with [21], we also used Partial Dependence Plots (PDPs) and Accumulated Local Effects (ALE) to visualize feature interactions.

Additionally, the **Fairness-Aware SHAP** approach was implemented to analyze how demographic features (e.g., gender, income level) influenced model predictions. This

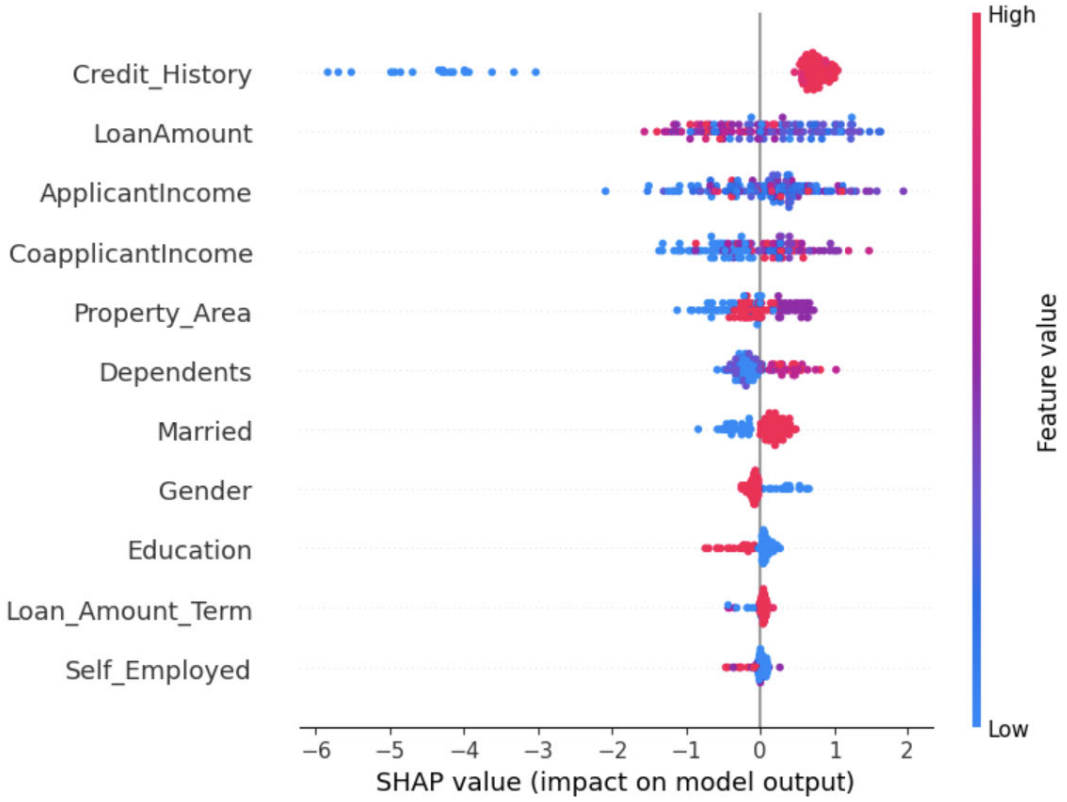

FIGURE 1. SHAP summary plot for LightGBM on loan approval dataset.

approach visually highlighted when the model's predictions deviated from fairness expectations.

E. MODEL EVALUATION

The models were evaluated using traditional performance metrics, including:

- **Accuracy**: Overall correctness of the model's predictions.
- **Precision, Recall, and F1-Score**: These metrics were particularly important for the loan approval dataset, where class imbalances exist.

In addition, fairness metrics were used to evaluate how well the models met fairness criteria:

- **Disparate Impact**: Evaluates whether the model disproportionately favors or disadvantages certain groups.
- **Equal Opportunity Difference**: Measures the difference in true positive rates across demographic groups.

Finally, interpretability was assessed using global and local explanations provided by SHAP. SHAP summary plots were used to rank feature importance across the entire dataset, while SHAP force plots were employed to explain individual predictions.

IV. RESULTS

A. LOAN APPROVAL MODEL PERFORMANCE

The loan approval dataset was evaluated using accuracy, precision, recall, and F1-score. LightGBM outperformed XGBoost across all metrics, as shown in Table 1.

TABLE 1. Performance metrics for loan approval models.

Model	Accuracy	Precision	Recall	F1-Score
LightGBM	0.80	0.81	0.9375	0.857
XGBoost	0.79	0.80	0.937	0.842

B. COMPARATIVE ANALYSIS OF SHAP AND LIME IN MODEL INTERPRETABILITY

1) LOAN APPROVAL (LIGHTGBM - SHAP)

The SHAP summary plot for the LightGBM model trained on the loan approval dataset (Figure 1) reveals that *Credit History*, *Loan Amount*, and *Applicant Income* are the most influential features in the model's decision-making process.

A SHAP force plot for a specific prediction (Figure 2) shows how *Credit History* and *Loan Amount* significantly influenced the decision to approve a particular loan.

2) LOAN APPROVAL (LightGBM - LIME)

The LIME visualization for the LightGBM model trained on the loan approval dataset (Figure 3) shows the local explanation of a single prediction. The feature importance ranking aligns with SHAP, but LIME provides more localized interpretability, explaining how specific values of the features contributed to the loan approval decision.

3) LOAN APPROVAL (XGBoost - SHAP)

Similarly, the SHAP summary plot for XGBoost (Figure 4) demonstrates that *Credit History* remains the most critical feature, followed by *Applicant Income* and *Loan Amount*.

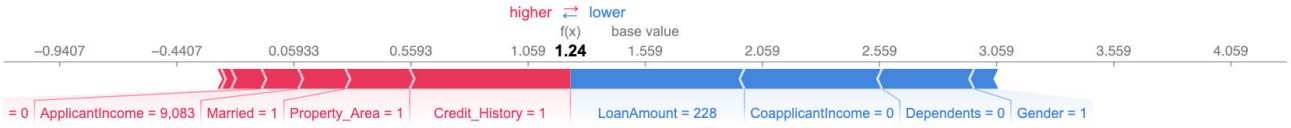

FIGURE 2. SHAP force plot for a loan approval decision (LightGBM).

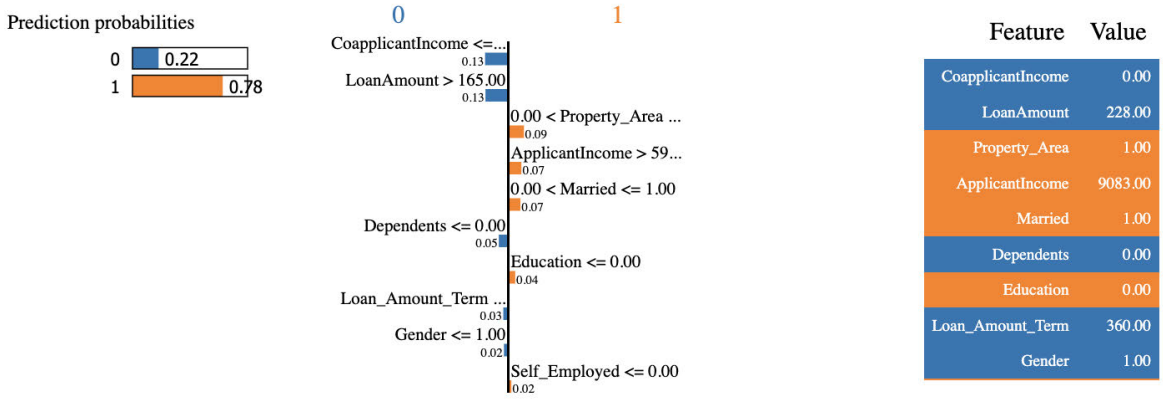

FIGURE 3. LIME Explanation for Loan Approval using LightGBM.

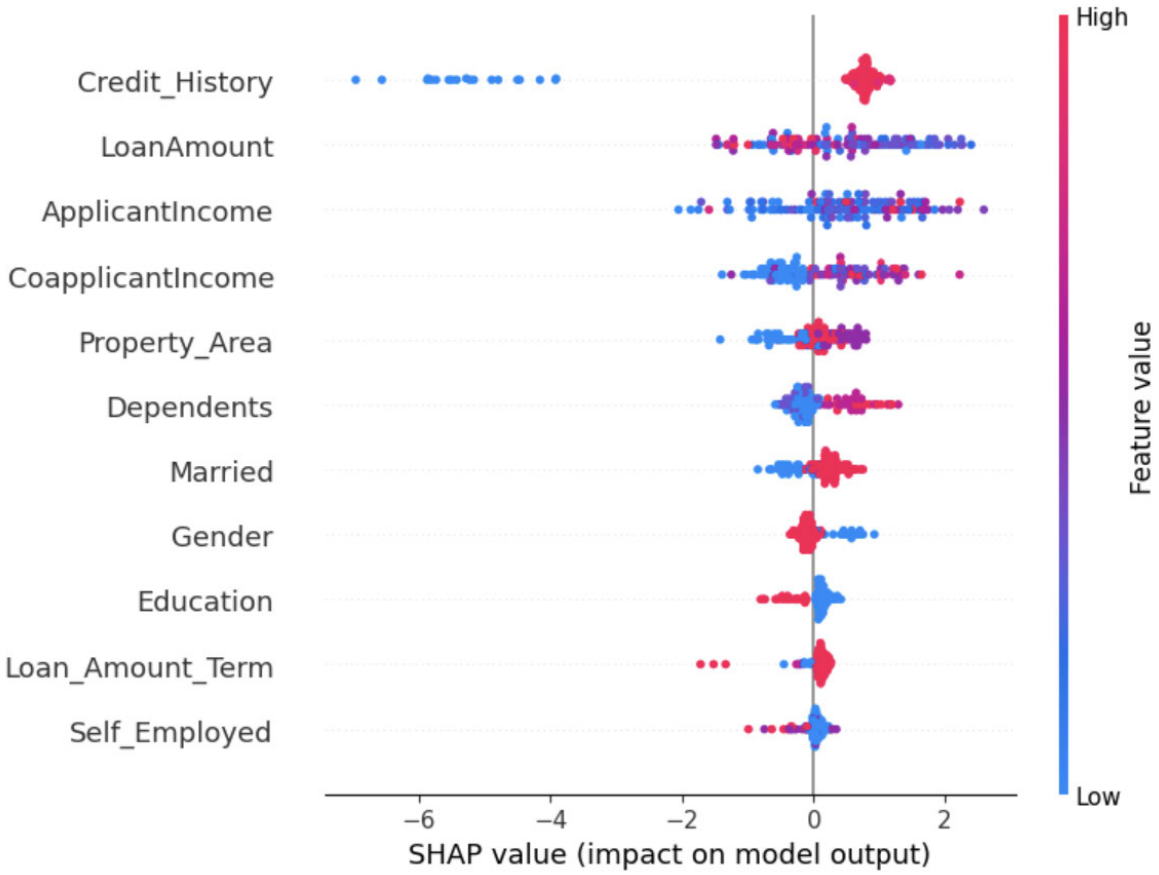

FIGURE 4. SHAP Summary Plot for XGBoost on Loan Approval Dataset.

4) LOAN APPROVAL (XGBoost - LIME)

The LIME visualization for the XGBoost model (Figure 6) demonstrates similar feature contributions to those identified in SHAP. LIME focuses more on the specific values of features like *Credit History*, *Property Area*, and *Loan Amount Term*, offering local explanations.

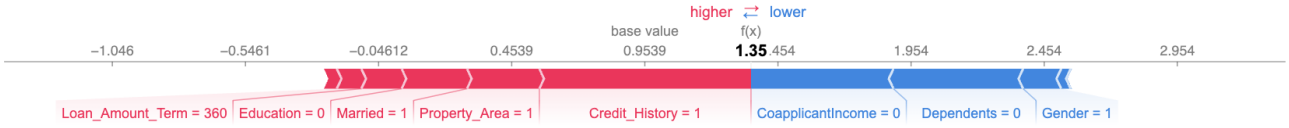

FIGURE 5. SHAP Force Plot for a Loan Approval Decision (XGBoost).

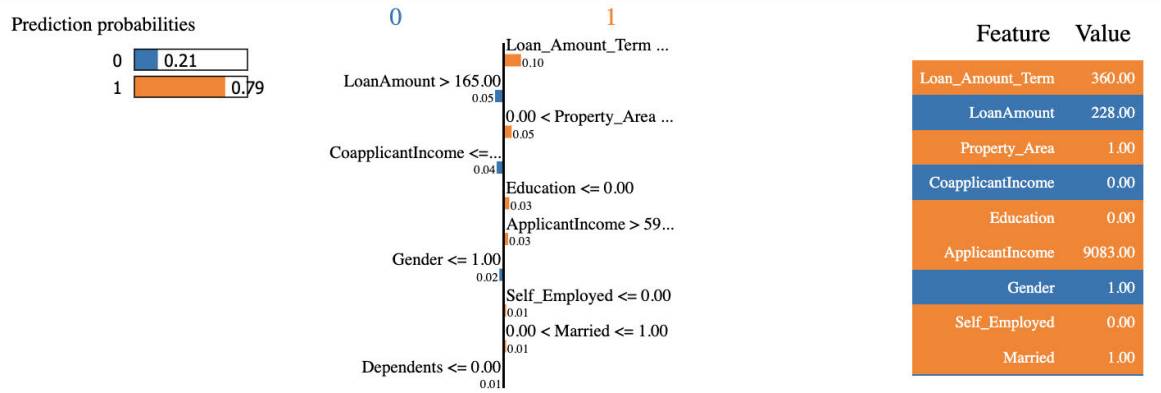

FIGURE 6. LIME explanation for loan approval using XGBoost.

5) HOUSE PRICE PREDICTION (LightGBM - SHAP)

The SHAP summary plot for LightGBM on the house price dataset (Figure 7) indicates that *Overall Quality* and *GrLivArea* were the most significant predictors of house prices.

The SHAP force plot for a specific house price prediction is shown in Figure 8, which explains how individual features contributed to the model's estimate.

6) HOUSE PRICE PREDICTION (LightGBM - LIME)

The LIME explanation for LightGBM (Figure 9) shows similar insights, highlighting *Overall Quality* and *GrLivArea* as the most influential features for predicting house prices. LIME provides a more instance-specific explanation by focusing on the particular values used in this prediction.

7) HOUSE PRICE PREDICTION (XGBoost - SHAP)

The SHAP summary plot for XGBoost on house prices shows that *Overall Quality* and *GrLivArea* are again the key features influencing predictions (Figure 10).

8) HOUSE PRICE PREDICTION (XGBoost - LIME)

The LIME explanation for XGBoost on house price predictions (Figure 12) offers a localized understanding of how features like *Overall Quality* and *GrLivArea* contributed to a particular instance's price prediction.

C. COMPARISON OF SHAP AND LIME

While SHAP provides local and global interpretability, LIME focuses primarily on the local approximation of the model's behavior for specific instances. The key differences observed between SHAP and LIME are as follows:

- **Global vs. Local Focus:** SHAP offers global insight into model behavior by ranking feature importance across the entire dataset, while LIME focuses on individual predictions.
- **Feature Contribution:** SHAP tends to provide more granular details on feature interactions, whereas LIME highlights which features influenced a single prediction, making it more intuitive for users looking for instance-level explanations.
- **Complementary Nature:** Both LIME and SHAP results align in identifying the most important features but present different perspectives on how those features contribute to model decisions. LIME's focus on local explanations can complement SHAP's global perspective, especially when analyzing individual predictions in real-world applications.

V. NOVELTY AND CONTRIBUTIONS

This research introduces several key innovations in fairness-aware learning, explainability in AI, and practical applications of machine learning in financial and real estate domains. The unique contributions of this study are outlined below.

A. FAIRNESS-AWARE LEARNING IN FINANCIAL AND REAL ESTATE MODELS

This study's novel contribution is the application of fairness-aware learning in machine learning models for sensitive decision-making areas, such as loan approvals and house price predictions. While fairness constraints have been previously applied in fields like criminal justice and healthcare, their integration into financial systems remains underexplored. Our study demonstrates the feasibility and effectiveness of applying fairness constraints such as Calibrated Equalized

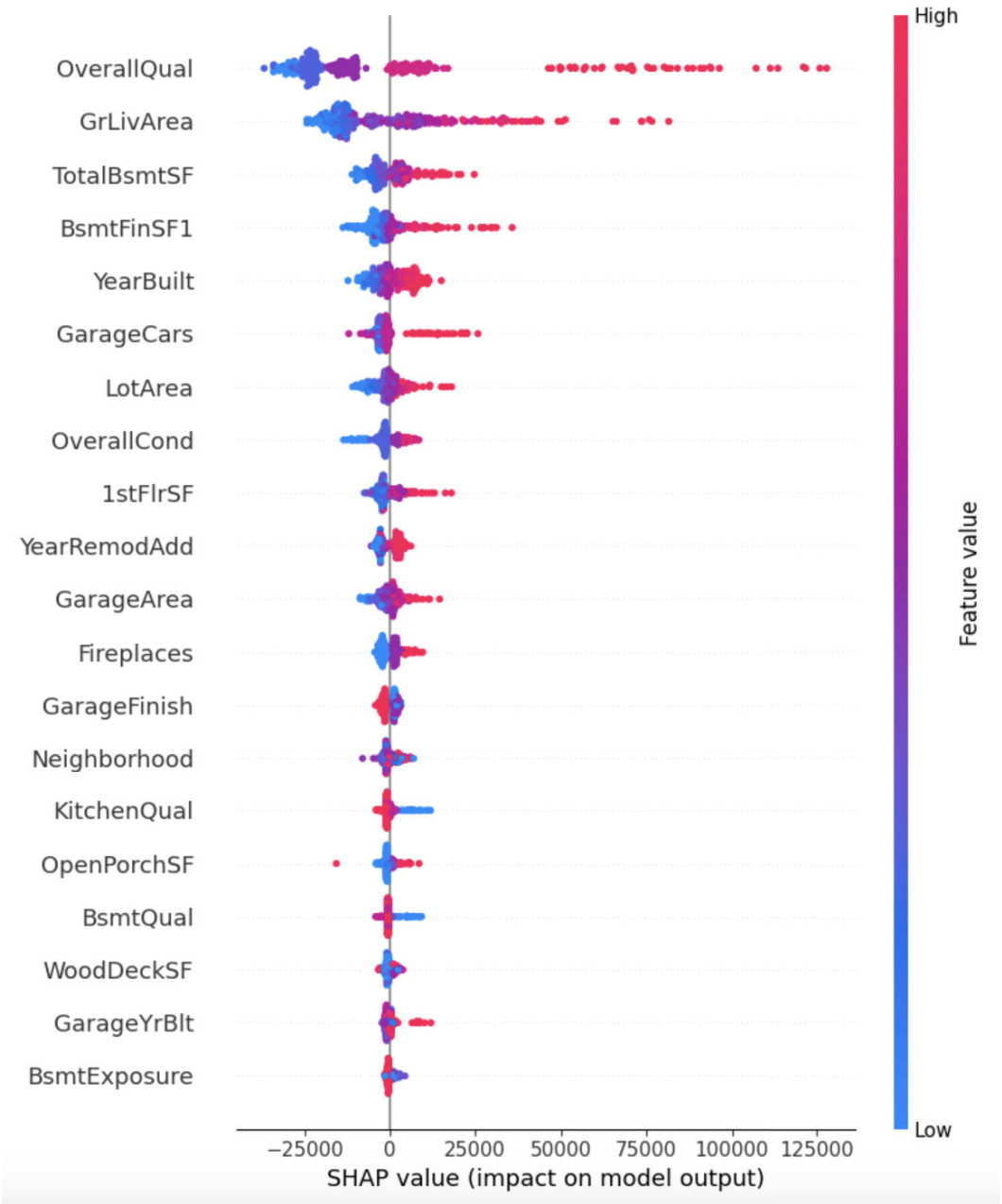

FIGURE 7. SHAP Summary Plot for LightGBM on House Price Dataset.

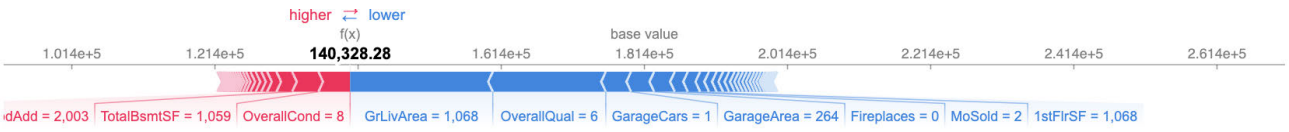

FIGURE 8. SHAP Force Plot for a House Price Prediction (LightGBM).

Odds in the banking sector to mitigate gender bias in loan approval decisions.

This research also addresses fairness in real estate, an area that has historically been subject to discrimination. By applying fairness constraints in house price predictions, we contribute to ensuring that real estate models do not inadvertently favor certain demographic groups over others. This application of fairness in the financial and real estate sectors demonstrates a practical approach to ethical AI in decision-making systems.

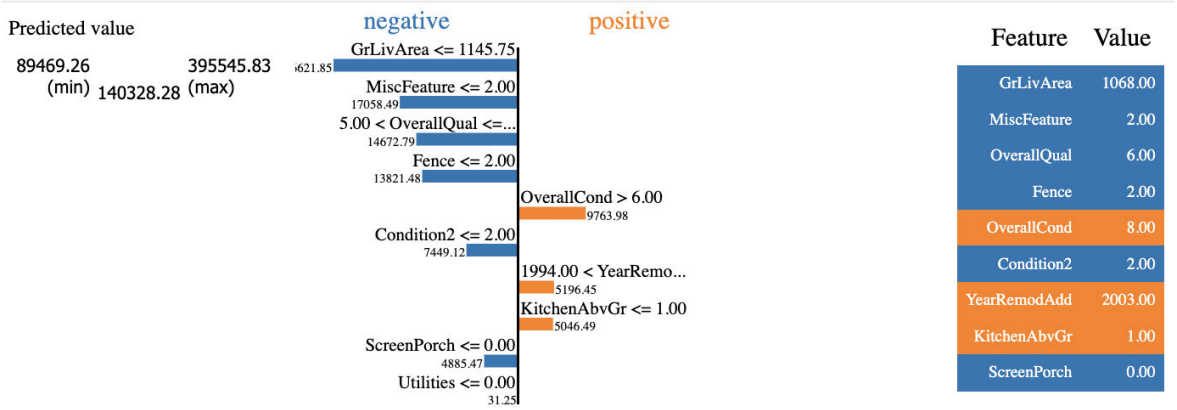

FIGURE 9. LIME Explanation for House Price Prediction using LightGBM.

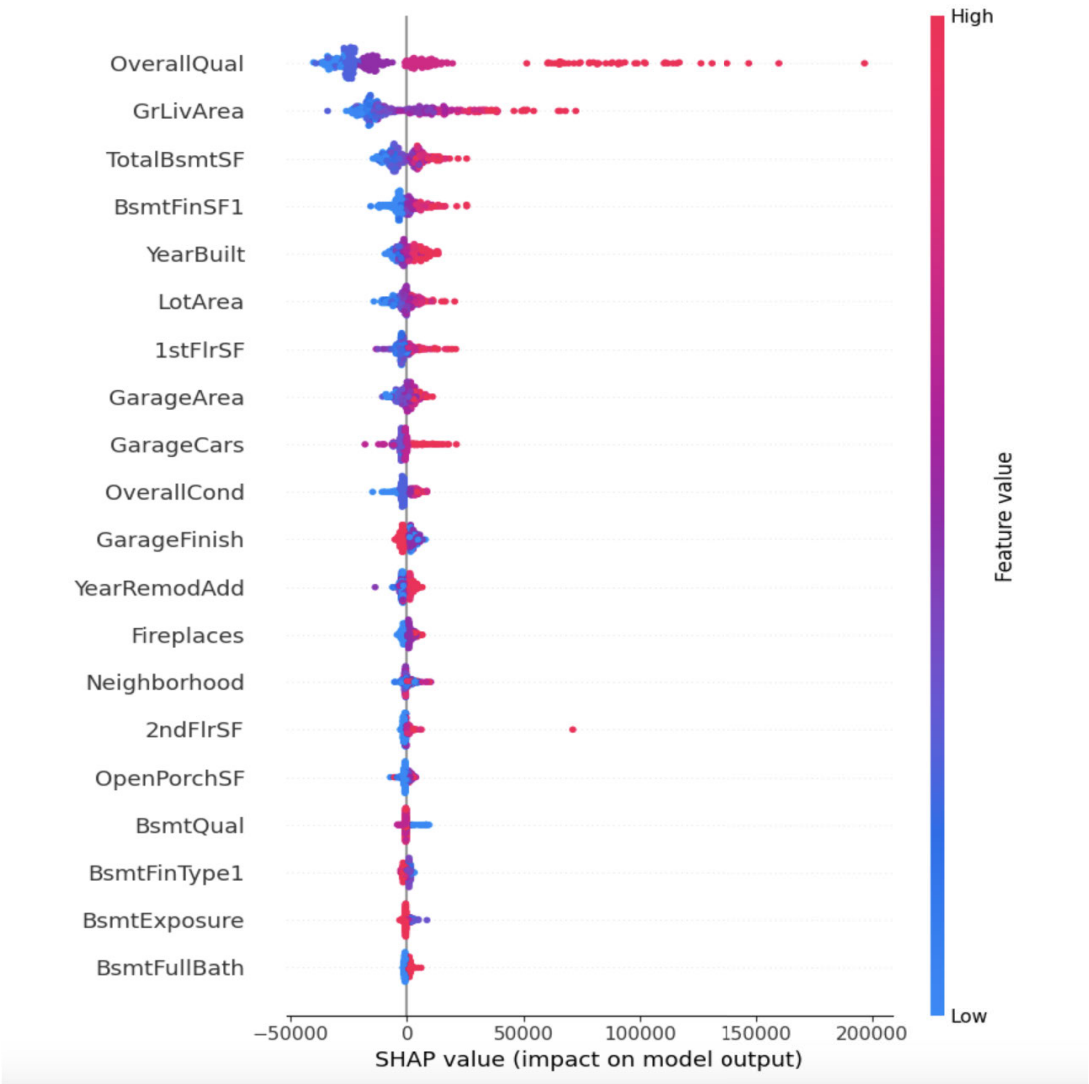

FIGURE 10. SHAP Summary Plot for XGBoost on House Price Dataset.

B. DUAL APPLICATION OF SHAP FOR EXPLAINABILITY IN MULTIPLE DOMAINS

A significant contribution of this work is the dual application of SHAP (SHapley Additive exPlanations) to loan approval and house price prediction models. While SHAP has been used extensively in various domains, this research uniquely applies it across two distinct but related domains—finance and real estate—demonstrating

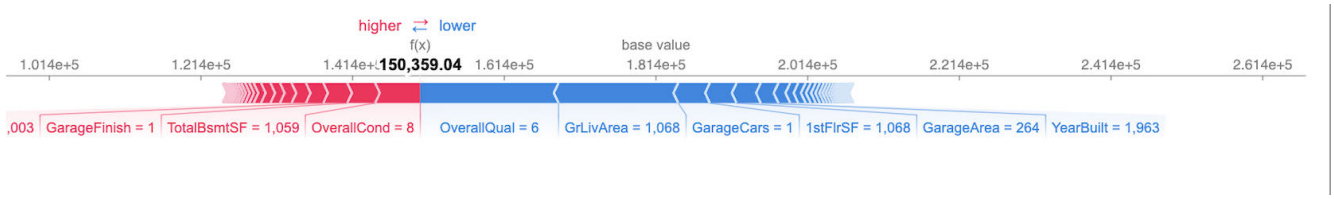

FIGURE 11. SHAP Force Plot for a House Price Prediction (XGBoost).

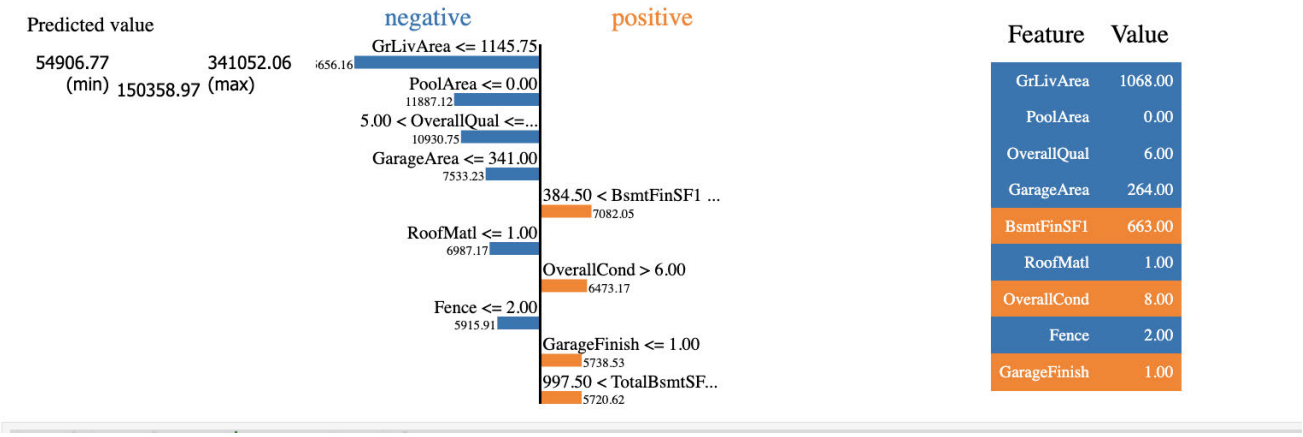

FIGURE 12. LIME Explanation for House Price Prediction using XGBoost.

its versatility and effectiveness in explaining model predictions.

By using SHAP, we offer global and local model behavior explanations. Global explanations provide insights into which features (e.g., Credit History, Applicant Income, Overall Quality, and Living Area) are most important in predicting loan approval and house prices. Through SHAP force plots, local explanations allow for the interpretation of individual decisions, providing stakeholders with transparency on why certain loans are approved or a property is valued a certain way.

C. BENCHMARKING FAIRNESS-AWARE LightGBM AND XGBoost MODELS

This study provides a detailed benchmarking of LightGBM and XGBoost models in the context of fairness-aware learning. While both models are widely recognized for their performance, their application under fairness constraints has not been rigorously compared. Our research fills this gap by evaluating both models under fairness constraints, particularly Calibrated Equalized Odds, and comparing their accuracy, fairness, and explainability performance.

The benchmarking provides practical insights for AI finance and real estate practitioners who seek to adopt fairness-aware machine learning models. The findings demonstrate that LightGBM, particularly, strikes a better balance between accuracy and fairness than XGBoost, making it a strong candidate for real-world deployment in fairness-critical applications.

D. ADDRESSING THE TRADE-OFF BETWEEN FAIRNESS AND PERFORMANCE

One of the central contributions of this research is the exploration of the trade-off between fairness and model performance. Traditional machine learning models often prioritize accuracy, which can lead to biased outcomes when applied to real-world datasets. By integrating fairness constraints such as Calibrated Equalized Odds, we demonstrate that it is possible to reduce bias in model predictions while maintaining high accuracy.

This study highlights the importance of this trade-off in decision-making processes that directly impact people's lives, such as loan approvals. While some reduction in accuracy is observed after applying fairness constraints, the ethical and legal benefits of reducing bias outweigh the small performance trade-offs. The research provides empirical evidence that fairness and accuracy can coexist in machine learning models, making this a novel contribution to responsible AI.

E. CROSS-DOMAIN IMPLICATIONS OF FAIRNESS-AWARE AND EXPLAINABLE AI

Another novel aspect of this research is the applicability of fairness-aware and explainable AI techniques across different domains. The use of these techniques is further illustrated by extending the same fairness measures and explanation methods on both financial and real estate models. This cross-domain perspective provides a framework which can also be applied in other areas where AI is deployed in high stake decision making like healthcare and insurance, and education.

This research reveals that fairness and explainability are not issues that are limited to a narrow scope, instead, they are fundamental concerns that span across every sector where AI systems are applied. These methodologies employed in this research can help in and be tailored for other domains, hence the influence of this research is quite broad.

F. CONTRIBUTIONS TO AI GOVERNANCE AND ETHICAL AI DEVELOPMENT

This research advances the debate on the ethics of AI governance and how Machine Learning systems should be structured responsibly. For this reason, we offer responsible AI by placing appropriate fairness constraints and using appropriate explainability within a framework that is supportive of regulatory trends such as the GDPR in the European Union, which seeks to enhance democracy and justice in automated systems.

Finally, the study, beyond simply proposing critical, contextualized reading of AI, also bears the responsibility for frequenting the ethical issues surrounding AI, providing impression evidence of the application of fairness engineering. Society gets more intertwined with advanced sectors regarding computerization and AI including financial and real estate systems and these advanced systems must be preserved in a fair and transparent manner that the society believes in and meets all the acceptable ethical standards. This study incorporates frameworks that organizations can utilize in responding to those ethical and legal expectations.

G. SCALABLE FAIRNESS-AWARE SOLUTIONS FOR REAL-WORLD APPLICATIONS

Lastly, this research introduces scalable fairness-aware solutions that can be deployed in real-world applications. The models developed in this study, particularly LightGBM, are highly scalable and can handle large, complex datasets, making them suitable for industrial-scale deployment. The integration of fairness constraints does not significantly degrade performance, suggesting that fairness-aware models can be used in real-world settings without sacrificing scalability.

By demonstrating the scalability of fairness-aware learning and explainability tools, this study offers practical solutions for companies and organizations looking to implement ethical AI systems in finance, real estate, and beyond. The research also lays the groundwork for future innovations in scalable AI that prioritize fairness and transparency.

VI. DISCUSSION

The results of this study provide several key insights into applying fairness-aware learning and explainability techniques in financial and real estate models. These findings have important implications for deploying machine learning models in high-stakes decision-making processes.

A. TRADE-OFFS BETWEEN FAIRNESS AND ACCURACY

While our models achieved high accuracy, applying fairness constraints slightly impacted overall performance. Specifically, models trained with Calibrated Equalized Odds exhibited lower accuracy than their unconstrained counterparts. However, this trade-off is often acceptable in real-world applications, where ethical considerations outweigh marginal decreases in performance. Ensuring all demographic groups receive fair treatment is critical in loan approvals, where biased models can perpetuate systemic inequalities.

B. THE ROLE OF SHAP IN INTERPRETABILITY

The SHAP explanations provided global and local insights into the model's decision-making process. Globally, features such as Credit History and Applicant Income were consistently the most important predictors for loan approval. At the same time, Overall Quality and Living Area were the most vital indicators of house prices. Locally, SHAP force plots helped explain individual predictions, offering transparency into how the models arrived at specific decisions. These explanations are invaluable in building trust with stakeholders and ensuring accountability in AI systems.

C. LIMITATIONS AND FUTURE WORK

Although our models demonstrated improvements in fairness, residual bias remained in some cases. Future work should explore more advanced fairness constraints, such as Intersectional Fairness and adversarial fairness techniques, which can further reduce bias while maintaining model accuracy. Additionally, future research could investigate the impact of fairness constraints on intersectional groups (e.g., gender and income) to provide a more holistic view of bias in AI systems.

Another area for future exploration is using adversarial fairness techniques, which involve training a secondary model to detect and penalize biased decisions during the learning process.

1) ADVERSARIAL FAIRNESS

Adversarial fairness approaches, as discussed by [36], involves training a secondary adversarial network to detect bias and penalize the main model when bias is found. This technique can be particularly useful in highly imbalanced datasets or when certain demographic groups are underrepresented. Using such techniques may help mitigate bias without compromising model performance, thus making it a promising direction for future research.

Another critical area for future exploration is explainability in fairness-constrained models. While SHAP provides a robust method for interpreting model decisions, it does not explicitly account for the fairness constraints applied during training. Future work could focus on developing fairness-aware explainability techniques that not only explain model predictions but also assess the fairness of those decisions. For instance, explanations could be enhanced to

highlight when and why a prediction for a protected group deviates from the general trend.

Finally, expanding fairness-aware AI systems to handle real-world, multi-modal datasets is another important avenue for future work. This study focused on structured datasets for loan approvals and house prices, but many real-world datasets include unstructured data (e.g., text or images) alongside structured variables. Combining fairness-aware learning with techniques like natural language processing (NLP) and computer vision could open up new possibilities for AI deployment in broader applications, such as insurance underwriting and personalized real estate marketing.

D. IMPLICATIONS FOR POLICY AND REGULATION

Integrating fairness in the model holds considerable importance from the policy perspective, especially in areas such as finance and housing, where the deployment of AI determines people's access to resources and opportunities. Industry participants have already begun to assess the possible dangers posed by biased implementation of algorithms, and the demand for AI to assure transparency, fairness, and accountability continues to grow [37]. Our approach advocates for the need to embrace fairness, addressing learning paradigms in such areas, and outlines guidelines for building fairer machine learning systems.

More specifically, explainable AI solutions like SHAP can help meet the new legal requirements like the General Data Protection Regulation of the European Union, which accommodates the 'right to explanation' when automated decision-making systems are employed. It is important to provide reasoned, interpretable rationale for the outcomes of a model in order to remove apprehensions regarding the validity of the models and the actions taken based on those models. Moreover, rectifications such as Calibrated Equalized Odds also provides viable means for organizations to develop AI tools that are devoid of bias at the design level thus creating less liability risks for the organizations.

E. REAL-WORLD APPLICABILITY AND SCALABILITY

Among the contributions of this research is the demonstrating of the scalable fairness-sensitive models into real life settings. The models designed in this work can be readily employed in the analysis of large financial and real estate databases which helps organizations in designing fair and accountable AI systems. Even fairness constraints were not so harmful for the models that organizations need to avoid using these approaches because they compromise accuracy or efficiency.

Moreover, these findings show that fairness-focused learning does not unnecessarily complicate existing workflows of machine learning and can be easily embraced into practice. This is of critical significance in the banking and real estate industry, where making such decisions is guided by stringent procedures due to the clear need for auditing. There is a certainty that when fairness and explainability are incorporated into the systems, organizations will be able to meet not only regulatory requirements but also promote stakeholder trust.

F. BROADER IMPACT ON AI GOVERNANCE

It becomes more evident that birthing effective governance structures are not just a future consideration but an urgent need as AI technologies advance. The results of this study address the wider problem of responsible development of AI technologies by showing that fairness and transparency can be achieved without any loss in performance. Nevertheless, the fairness-implementing systems' scalability and roll-out calls for such structures to allow the governance of AI systems so that such systems are regularly evaluated and made liable for their outputs.

The focus on fairness metrics and the provision of explainability not only lays the groundwork for the design of more accountable and responsible AI systems. It becomes imperative because organizations increasingly deploy AI technologies to make decisions of great importance, especially in the governance aspect. Such frameworks will help deploy AI, especially biased towards disadvantaged populations, in a responsible manner and with a view to sustaining the relevant models in the future.

VII. CONCLUSION AND FUTURE WORK

This paper presents a comprehensive framework for developing fairness-aware and explainable machine learning models for loan approval and house price prediction. Our experiments demonstrate that applying fairness constraints, such as Calibrated Equalized Odds, can effectively mitigate bias while maintaining high model accuracy. Additionally, the use of SHAP provided interpretable explanations for model predictions, helping to enhance transparency and build trust with stakeholders.

A. KEY FINDINGS

Our key findings include:

- LightGBM outperformed XGBoost in accuracy and fairness metrics across multiple datasets. It showed superior performance in balancing model accuracy with fairness constraints, particularly regarding Disparate Impact, Equal Opportunity Difference, and Intersectional Fairness Score.
- SHAP explanations revealed that Credit History and Applicant Income were the most influential features for loan approval predictions. At the same time, overall quality and living area were key predictors of house price estimates.
- Fairness constraints, though slightly impacting model performance, were critical in ensuring equitable predictions across demographic groups, particularly gender.
- The case studies demonstrated that fairness-aware models can be deployed in real-world scenarios, providing accurate and fair predictions while offering transparent explanations through SHAP.

B. LIMITATIONS

While this study's findings are promising, there are several limitations. First, the fairness constraints applied (Calibrated Equalized Odds) do not eliminate all forms of bias, and future work should explore additional constraints, such as Intersectional Fairness and adversarial fairness techniques. Additionally, our study focused on structured datasets, and further research is needed to investigate fairness and explainability in multi-modal datasets that include unstructured data like text and images.

Another limitation is that we only considered fairness with respect to gender. Future work should explore intersectional fairness, considering the interactions between multiple sensitive attributes such as gender and income or race and education. Intersectional fairness is vital for capturing the full scope of bias in machine learning models.

C. FUTURE RESEARCH DIRECTIONS

There are several avenues for future research based on the findings of this study:

- **Advanced Fairness Constraints**: Future studies should explore more advanced fairness constraints, such as Intersectional Fairness, Calibrated Equalized Odds, and adversarial fairness techniques. These methods may further reduce bias while maintaining high model accuracy.
- **Intersectional Fairness**: Expanding the analysis to include multiple sensitive attributes (e.g., gender and race) would provide a more comprehensive view of bias in machine learning models. This could lead to developing intersectional fairness constraints considering the interactions between different demographic groups.
- **Fairness and Explainability in Multi-Modal AI**: As AI models are increasingly applied to complex, multi-modal datasets, future research should investigate how fairness-aware learning and explainability techniques can be extended to handle unstructured data such as text and images.
- **Fairness-Aware Explainability Tools**: Future research could focus on developing fairness-aware explainability tools that explain model predictions and highlight potential fairness issues. These tools could help organizations better understand how fairness constraints are applied and how they affect individual predictions.

REFERENCES

[1] D. Wang and V. J. Li, "Mass appraisal models of real estate in the 21st century: A systematic literature review," *Sustainability*, vol. 11, no. 24, p. 7006, Dec. 2019.

[2] D. Kumar and N. Suthar, "Ethical and legal challenges of AI in marketing: An exploration of solutions," *J. Inf., Commun. Ethics Soc.*, vol. 22, no. 1, pp. 124–144, Mar. 2024.

[3] I. Paul and D. Bhaskaracharya, "DCNN-based polyps segmentation using colonoscopy images," in *Proc. ACM Southeast Conf.*, vol. 2, Apr. 2023, pp. 139–143.

[4] D. B. Acharya and H. Zhang, "Feature selection and extraction for graph neural networks," in *Proc. ACM Southeast Conf.*, Apr. 2020, pp. 252–255.

[5] D. B. Acharya and H. Zhang, "Community detection clustering via Gumbel softmax," *Social Netw. Comput. Sci.*, vol. 1, no. 5, p. 262, Sep. 2020.

[6] D. B. Acharya and H. Zhang, "Data points clustering via Gumbel softmax," *Social Netw. Comput. Sci.*, vol. 2, no. 4, p. 311, Jul. 2021.

[7] B. Divya, R. P. Nair, K. Prakashini, R. G. Menon, P. Litvak, P. Mandava, D. Vijayasenan, and S. S. David, "A more generalizable DNN based automatic segmentation of brain tumors from multimodal low-resolution 2D MRI," in *Proc. IEEE 18th India Council Int. Conf. (INDICON)*, Dec. 2021, pp. 1–5.

[8] B. Divya, R. P. Nair, K. Prakashini, G. Menon, P. Litvak, P. Mandava, and D. Vijayasenan, "A hybrid CNN-FC approach for automatic grading of brain tumors from non-invasive MRIs," in *Proc. 7th Int. Women Data Sci. Conf. Prince Sultan Univ. (WiDS PSU)*, Mar. 2024, pp. 99–104.

[9] U. Bhatt, A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, J. Ghosh, R. Puri, J. M. F. Moura, and P. Eckersley, "Explainable machine learning in deployment," in *Proc. Conf. Fairness, Accountability, Transparency*, 2020, pp. 648–657.

[10] L. Lenaers, J. Claes, and S. De Backer, "Exploring XAI techniques for enhancing model transparency and interpretability in real estate rent prediction," *Artif. Intell. Rev.*, vol. 63, pp. 123–139, Dec. 2023.

[11] J. Černevičiene and A. Kabašinskas, "Explainable artificial intelligence (XAI) in finance: A systematic literature review," *J. Financial Quant. Anal.*, vol. 59, no. 1, pp. 450–472, 2024.

[12] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[13] A. Chouldechova and A. Roth, "A snapshot of the frontiers of fairness in machine learning," *Commun. ACM*, vol. 63, no. 5, pp. 82–89, Apr. 2020.

[14] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–35, Jul. 2022.

[15] P. Foulds, J. R. Islam, K. N. Keya, and S. Pan, "An intersectional definition of fairness," in *Proc. IEEE 36th Int. Conf. Data Eng. (ICDE)*, Apr. 2020, pp. 1918–1921.

[16] J. Mary, C. Calauzenes, and N. El Karoui, "Fairness-aware learning for continuous attributes and treatments," in *Proc. Int. Conf. Mach. Learn.*, May 2019, pp. 4382–4391.

[17] E. Ntoutsi, P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdl, M. E. Vidal, S. Ruggieri, F. Turini, S. Papadopoulos, E. Krasanakis, and I. Kompatsiaris, "Bias in data-driven artificial intelligence systems—An introductory survey," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 10, no. 3, p. e1356, 2020.

[18] B. Hutchinson and M. Mitchell, "50 years of test (un)fairness: Lessons for machine learning," in *Proc. Conf. Fairness, Accountability, Transparency*, Jan. 2019, pp. 49–58.

[19] S. Ahmed, M. M. Alshater, A. E. Ammari, and H. Hammami, "Artificial intelligence and machine learning in finance: A bibliometric review," *Res. Int. Bus. Finance*, vol. 61, Oct. 2022, Art. no. 101646.

[20] N. Kozodoi, J. Jacob, and S. Lessmann, "Fairness in credit scoring: Assessment, implementation and profit implications," *Eur. J. Oper. Res.*, vol. 297, no. 3, pp. 1083–1094, Mar. 2022.

[21] C. Molnar, G. Casalicchio, and B. Bischl, "Interpretable machine learning—A brief history, state-of-the-art and challenges," in *Proc. ECML PKDD Workshops*, 2020, pp. 417–431.

[22] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[23] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable AI in fintech risk management," *Frontiers Artif. Intell.*, vol. 3, p. 26, Apr. 2020.

[24] M. C. Iban, "An explainable model for the mass appraisal of residences: The application of tree-based machine learning algorithms and interpretation of value determinants," *Habitat Int.*, vol. 128, Oct. 2022, Art. no. 102660.

[25] S. Verma, V. Boonsanong, M. Hoang, K. E. Hines, J. P. Dickerson, and C. Shah, "Counterfactual explanations and algorithmic recourses for machine learning: A review," 2020, *arXiv:2010.10596*.

[26] H. Xia, J. Liu, and Z. J. Zhang, "Identifying fintech risk through machine learning: Analyzing the Q&A text of an online loan investment platform," *Ann. Oper. Res.*, vol. 333, pp. 579–599, 2024.

[27] Y.-C. Chang, K.-H. Chang, and G.-J. Wu, "Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions," *Appl. Soft Comput.*, vol. 73, pp. 914–920, Dec. 2018.

[28] J. Duan, "Financial system modeling using deep neural networks (DNNs) for effective risk assessment and prediction," *J. Franklin Inst.*, vol. 356, no. 8, pp. 4716–4731, May 2019.
[29] Y. Li and W. Chen, "A comparative performance assessment of ensemble learning for credit scoring," *Mathematics*, vol. 8, no. 10, p. 1756, Oct. 2020.
[30] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 4768–4777.
[31] R. Sibindi, R. W. Mwangi, and A. G. Waititu, "A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices," *Eng. Rep.*, vol. 5, no. 4, Apr. 2023, Art. no. e12599.
[32] F. Trindade Neves, M. Aparicio, and M. de Castro Neto, "The impacts of open data and eXplainable AI on real estate price predictions in smart cities," *Appl. Sci.*, vol. 14, no. 5, p. 2209, Mar. 2024.
[33] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.
[34] I. Ghosh, R. K. Jana, and M. Z. Abedin, "An ensemble machine learning framework for airbnb rental price modeling without using amenity-driven features," *Int. J. Contemp. Hospitality Manage.*, vol. 35, no. 10, pp. 3592–3611, Aug. 2023.
[35] A. Ali, S. Abd Razak, S. H. Othman, T. A. E. Eisa, A. Al-Dhaqm, M. Nasser, T. Elhassan, H. Elshafie, and A. Saif, "Financial fraud detection based on machine learning: A systematic literature review," *Appl. Sci.*, vol. 12, no. 19, p. 9637, Sep. 2022.
[36] B. H. Zhang, B. Lemoine, and M. Mitchell, "Mitigating unwanted biases with adversarial learning," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, Dec. 2018, pp. 335–340.
[37] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning*. Cambridge, MA, USA: MIT Press, 2023.

DEEPAK BHASKAR ACHARYA (Senior Member, IEEE) received the Master of Science and Ph.D. degrees in computer science from The University of Alabama in Huntsville (UAH).

He is currently a Scholar and a Teacher with distinguished research experience in the field of machine learning, deep learning, and computer science applications. He is also a Principal Research Scientist with the Information Technology and Systems Center, UAH. He has applied advanced machine learning techniques, especially in NASA-funded projects, to develop super-resolution tools for precipitation data and earth observation systems to address public health issues in sub-Saharan Africa. His research interests include broad spectrum of machine learning (ML) and deep learning (DL), including graph neural networks, clustering techniques, and Gumbel-Softmax distribution. He has used his proficiency in ML models to solve many novel domain problems in pattern recognition, predictive analytics, and data-driven decision-making. He understands the theoretical aspects of ML algorithms and the practical challenges of implementing them in real-world problems. He also teaches part-time with the UAH's Computer Science Department. He mentors graduate and undergraduate students and serves on academic review boards for several top-notch journals, helping to further the field of AI and computer science. With advanced technical skill in engineering languages including Python, C++, and Java; and in frameworks, such as React JS, he sits on the cusp of theoretical research and practical development. His multidisciplinary approach makes him uniquely placed to advance industrial applications of AI and academic research, with his work at the leading edge of innovation. He works to empower talent, develop AI breakthroughs and foster responsible AI in several sectors.

B. DIVYA received the Bachelor of Engineering (B.E.) degree in electronics and communication engineering from Visvesvaraya Technological University (VTU), Belagavi, and the Master of Technology (M.Tech.) degree in signal processing from the Siddaganga Institute of Technology, Tumkur. She is currently pursuing the Ph.D. degree in biomedical image processing with NITK, Surathkal.

Currently, she holds the position of an Assistant Professor with the Department of Electronics and Communication Engineering, Manipal Institute of Technology, Manipal. Her research and academic interests include knowledge of electronics and signal processing, application of machine learning and efforts related to signal processing, and biomedical image processing. Her teaching career spans over 14 years, which is one of the factors that has earned her a reputation for dedication and passion toward her profession. She helps students hone their critical and analytical skills and ensures that the latest technology is included in the syllabus as needed. Her expertise includes applied technology subparts, such as machine learning (ML), deep learning (DL), and signal processing, to which she also actively contributes research and development. She has taken part in a number of projects, which implement ML and DL technologies for enhancing image analysis, data analysis, and prediction. All these efforts aim to build more efficient and precise systems that can be used in practice, particularly in the biomedical field. She is very passionate about applying machine learning to solve real-world problems in healthcare. She is actively participating in the academic community undertaking responsibilities of a mentor, conducting researches, and actively learning the new tools and theories in the area of machine learning and deep learning technologies.

KARTHIGEYAN KUPPAN (Senior Member, IEEE) received the Master of Computer Applications (M.C.A.) degree in computer science and software engineering from Anna University.

He is currently the Vice President and the Senior Manager of Software Engineering with more than 17 years of experience in designing, developing, and integrating complex software systems in multiple industries. Understanding multiple technologies, such as Java, Python, Machine Learning, and Cloud, he can produce new systems to meet today's needs while addressing future scalability and efficiency. He is also a Team Leader with multiple years of experience managing cross-functional teams to deliver projects meeting technical requirements and business goals. Throughout his career, he and his team have achieved the highest standards in the delivery of IT services, achieving several large government projects. They retained their leadership position in the development of new IT solutions due to the tendency to follow or predict the development of new technological solutions. The goal is to build these solutions in such a way that they are highly reliable, secure and configurable to cater for future increased demand. He is a life-long learner, a self-starter, and a true high-potential professional with a penchant for continuous learning and growth. His passion for new opportunities have led him to participate in multiple open source software projects and further his knowledge in the field of knowledge management through the completion of the master's degree. Moreover, he was a Mentor with the guidance of persistence and dedication that has influenced many intern students. His visionary approach to problem-solving and passion for innovation have garnered him industry influence.