Grenze International Journal of Engineering and Technology, June Issue

Enhancing Model Interpretability using Local Interpretable Model-Agnostic Explanations (Lime): Insights from Predictive Analysis on Housing Market Dynamics

Naga Siva Jyothi Kompalli[1], Medavarapu Swethan Rao[2], Pilli Uttej[3] and Kancharla Jayakanth Reddy[4]
[1-4]Department of IT, SNIST, Hyderabad, India
Email: jyothi.kompalli@gmail.com, swethanrao1671@gmail.com, yahodiuttej7@gmail.com, jayakanthreddykancharla@gmail.com

***Abstract*— This paper examines the interpretability of Random Forest Regression models in real estate using Local Interpretable Model-agnostic Explanations (LIME). Analysing LIME explanations for selected instances, we uncover crucial insights into the factors influencing house price predictions. Notably, features such as socioeconomic indicators and environmental metrics significantly impact predicted prices. We highlight the importance of model interpretability for transparency and trust in complex domains like real estate. By employing techniques like LIME, researchers and practitioners can make more informed decisions based on underlying factors. This study contributes to advancing model interpretation, suggesting future research directions for applying LIME in other domains and exploring its effectiveness in diverse predictive models.**

***Index Terms*— Local Interpretable Model-agnostic Explanations (LIME), Black Box Machine Learning Models, Model Interpretability, Prediction Rationales, Machine Learning Explainability**

I. INTRODUCTION

A. Background on Model Interpretability

Model interpretability [1] has emerged as a critical aspect in the realm of machine learning and artificial intelligence (AI). With the proliferation of complex black box models, understanding the rationale behind their predictions has become increasingly challenging. Interpretability refers to the ability to elucidate how a model arrives at its decisions, providing insights into the underlying patterns and relationships within the data. Traditional linear models offer inherent interpretability [2] due to their transparent nature, but the advent of more sophisticated models, such as neural networks and ensemble methods, has ushered in an era of opaque models with superior predictive performance but limited interpretability.

B. Importance of Understanding Model Predictions

Understanding the predictions of machine learning models is crucial for various reasons, spanning both practical and ethical considerations. From a practical standpoint, interpretable models foster trust and confidence among end-users, stakeholders, and regulatory bodies, enabling informed decision-making and risk assessment.

Grenze ID: 01.GIJET.10.2.172

Moreover, interpretable models facilitate debugging, error analysis, and model refinement, thereby enhancing overall model robustness and reliability [3]. On an ethical level, transparent models help mitigate the potential risks of biased or discriminatory decisions, promoting fairness, accountability, and transparency in AI systems.

C. Overview of Local Interpretable Model-agnostic Explanations (LIME)

Local Interpretable Model-agnostic Explanations (LIME) offers a systematic framework [4] for interpreting the predictions of black box machine learning models [5]. Rooted in the principles of model agnosticism, LIME aims to provide local explanations for individual predictions by approximating the underlying decision boundaries of complex models. Unlike global interpretation methods, which seek to explain the entire model's behaviour, LIME focuses on generating explanations specific to a given instance of interest. By training interpretable surrogate models on locally perturbed data samples, LIME offers a granular understanding of prediction rationales without compromising model complexity or predictive performance.

D. Motivation for Using LIME in Model Interpretation

The motivation [6] for employing LIME in model interpretation stems from the inherent challenges posed by black box models, including neural networks, support vector machines, and random forests, among others. While these models excel in predictive accuracy, their opaque nature hinders understanding and trust among end-users. LIME addresses this challenge by providing interpretable explanations tailored to individual predictions, thereby bridging the gap between model complexity and interpretability. By elucidating the factors contributing to each prediction, LIME enhances transparency, accountability, and trust in machine learning systems, paving the way for responsible AI deployment across diverse applications and domains.

II. Literature Review

A. Existing Approaches to Model Interpretation

Model interpretation has been a longstanding research area [7] in machine learning, with various approaches developed to shed light on the inner workings of complex models. Traditional techniques, such as feature importance analysis, partial dependence plots, and decision trees, offer insights into model behaviour but are often limited in their ability to capture nonlinear relationships and interactions within the data. More sophisticated methods, including SHapley Additive explanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), and integrated gradients, have emerged to address these limitations by providing local and global explanations for black box models. SHAP leverages cooperative game theory to assign credit to each feature based on its contribution to the model's output, while LIME generates interpretable surrogate models to approximate local model behaviour. These methods offer versatile tools for interpreting a wide range of machine learning models, thereby enhancing transparency and trust in predictive systems.

B. Limitations of Traditional Interpretability Techniques

Despite their utility, traditional interpretability techniques suffer from several limitations when applied to modern machine learning models. Linear models, while inherently interpretable, may fail to capture complex nonlinear relationships present in high-dimensional datasets. Decision trees, although intuitive, are prone to overfitting and lack generalizability across different domains. Feature importance metrics, such as coefficients in linear models or variable importance in decision trees, provide limited insights into the direction and magnitude of feature contributions. Additionally, these techniques often struggle to explain individual predictions, focusing instead on model-level summaries. As a result, there is a growing need for more sophisticated interpretability methods capable of providing nuanced explanations for complex models on a local and global scale.

C. Evolution of LIME and Its Theoretical Foundations

Local Interpretable Model-agnostic Explanations (LIME) represents a significant advancement in the field of model interpretation [8], offering a principled approach to explain individual predictions of black box models. Founded on the principles of model agnosticism, LIME aims to approximate the behaviour of complex models using interpretable surrogate models trained on locally perturbed data samples. The theoretical foundation of LIME lies in the optimization of local surrogate models that minimize a loss function while constraining model complexity. By generating explanations specific to each prediction, LIME provides valuable insights into the factors driving model decisions [9], thereby enhancing transparency and trust in machine learning systems. The evolution of LIME underscores the ongoing efforts to bridge the gap between model complexity [10] and interpretability, paving the way for more responsible and accountable AI deployment in real-world applications.

III. Problem Statement

A. Lack of Transparency in Black Box Machine Learning Models

The proliferation of complex black box machine learning models, including deep neural networks, ensemble methods, and support vector machines, has led to a fundamental lack of transparency in model decision-making processes [11]. While these models often exhibit superior predictive performance, their opaque nature hinders understanding and trust among end-users. Stakeholders, regulators, and domain experts are unable to ascertain how these models arrive at their predictions, raising concerns regarding accountability, fairness, and potential biases. Without transparency, it becomes challenging to validate model behaviour, diagnose errors, and ensure ethical deployment of AI systems in real-world scenarios.

B. Challenges in Interpreting Model Predictions

Interpreting the predictions of black box machine learning models poses several challenges due to their inherent complexity and nonlinearity [12]. Traditional interpretability techniques, such as feature importance analysis and partial dependence plots, may offer insights at a global level but struggle to provide nuanced explanations for individual predictions. Moreover, these techniques often fail to capture intricate interactions and non-monotonic relationships present in high-dimensional datasets. As a result, there is a pressing need for more sophisticated interpretation methods capable of offering local explanations tailored to specific instances, thereby enhancing transparency and trust in predictive systems.

C. Need for Systematic Frameworks like LIME

The lack of transparency and challenges in interpreting model predictions underscore the need for systematic frameworks like Local Interpretable Model-agnostic Explanations (LIME). LIME offers a principled approach to model interpretation by approximating the behaviour of black box models using interpretable surrogate models [13]. By training these surrogate models on locally perturbed data samples, LIME provides localized explanations for individual predictions, thereby bridging the gap between model complexity and interpretability. The systematic nature of LIME enables stakeholders to gain insights into the factors driving model decisions, fostering transparency, accountability, and trust in machine learning systems. As such, there is a growing demand for the adoption of systematic interpretation frameworks like LIME to ensure the responsible and ethical deployment of AI technologies across diverse applications and domains.

IV. Objectives

A. Develop Python Implementation of LIME for Model Interpretation

The primary objective of this research is to develop a Python implementation of Local Interpretable Model-agnostic Explanations (LIME) for model interpretation [14]. This involves implementing the core algorithms and methodologies of LIME, including the generation of interpretable surrogate models and the approximation of local model behaviour. The Python implementation will adhere to best practices in software engineering and be designed to facilitate ease of use and extensibility for researchers and practitioners in the field of machine learning.

B. Demonstrate the Application of LIME in Interpreting Predictions from Random Forest Regressor

Another objective of this research is to demonstrate the practical application of LIME in interpreting predictions from a Random Forest Regressor model. By applying LIME to a real-world predictive model [15], such as the Random Forest Regressor, we aim to showcase the effectiveness of LIME in providing interpretable explanations for individual predictions. Through a series of case studies and experiments, we will elucidate the factors driving model decisions and highlight the insights gained from LIME explanations.

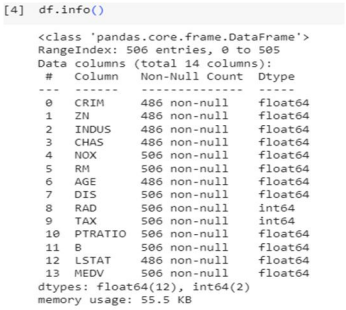
C. Evaluate the Performance of LIME in Providing Model Explanations

The final objective of this research is to evaluate the performance of LIME in providing model explanations. This involves conducting comprehensive experiments to assess the accuracy, stability, and fidelity of LIME explanations across different machine learning models and datasets. We will compare the interpretability of LIME explanations with existing interpretation methods and evaluate its effectiveness in capturing complex model behaviour. By quantitatively assessing the performance of LIME, we aim to validate its utility as a systematic framework for model interpretation and promote its adoption in both research and practical applications.

V. Methodology

A. Data Loading and Preprocessing

This phase involves loading the dataset and preprocessing it to ensure compatibility with the LIME framework and the Random Forest Regressor model. Data loading entails accessing the dataset from the appropriate source and importing it into the Python environment. Preprocessing steps may include handling missing values, encoding categorical variables, scaling numerical features, and splitting the data into training and testing sets. The goal is to prepare a clean and standardized dataset that can be used for model training and interpretation.



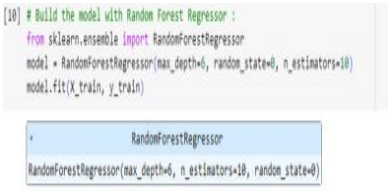
```
[4]  df.info()

     <class 'pandas.core.frame.DataFrame'>
     RangeIndex: 506 entries, 0 to 505
     Data columns (total 14 columns):
      #   Column   Non-Null Count  Dtype
     ---  ------   --------------  -----
      0   CRIM     486 non-null    float64
      1   ZN       486 non-null    float64
      2   INDUS    486 non-null    float64
      3   CHAS     486 non-null    float64
      4   NOX      506 non-null    float64
      5   RM       506 non-null    float64
      6   AGE      486 non-null    float64
      7   DIS      506 non-null    float64
      8   RAD      506 non-null    int64
      9   TAX      506 non-null    int64
      10  PTRATIO  506 non-null    float64
      11  B        506 non-null    float64
      12  LSTAT    486 non-null    float64
      13  MEDV     506 non-null    float64
     dtypes: float64(12), int64(2)
     memory usage: 55.5 KB
```

Figure1: Dataset Summary

B. Model Development with Random Forest Regressor

In this step, we develop a Random Forest Regressor model using the pre-processed dataset. The Random Forest Regressor is chosen as the predictive model due to its versatility and widespread use in regression tasks. The model is trained on the training dataset using the appropriate hyperparameters and optimization techniques. The trained model is then evaluated on the testing dataset to assess its performance in terms of predictive accuracy and generalization to unseen data.



```
[10] # Build the model with Random Forest Regressor :
     from sklearn.ensemble import RandomForestRegressor
     model = RandomForestRegressor(max_depth=6, random_state=0, n_estimators=10)
     model.fit(X_train, y_train)

     RandomForestRegressor
     RandomForestRegressor(max_depth=6, n_estimators=10, random_state=0)
```

Figure2: Random Forest Regressor

C. Explanation of Model Predictions using LIME

LIME is applied to interpret the predictions of the trained Random Forest Regressor model. This involves creating an explainer object using the LIME framework and specifying the relevant parameters, such as the number of features to be considered and the distance metric for perturbing the data. LIME explanations are generated for a subset of instances from the testing dataset, providing local interpretations for individual predictions. The explanations highlight the contribution of each feature to the model's predictions and offer insights into the factors driving the model's decisions.

D. Interpretation of LIME Results

The LIME results are interpreted to gain a deeper understanding of the Random Forest Regressor model and its predictions. This involves analysing the generated explanations and identifying patterns or trends in the feature contributions across different instances. The interpretations may involve comparing LIME explanations with domain knowledge or conducting sensitivity analyses to assess the robustness of the explanations. The goal is to extract actionable insights from the LIME results and use them to enhance understanding, trust, and transparency in the predictive model.

VI. RESULTS

A. Evaluation of Model Performance

The performance of the Random Forest Regressor model is assessed using standard evaluation metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2). These metrics offer insights into the model's predictive accuracy and its ability to generalize to unseen data. Additionally,

performance analysis may involve visualizations such as scatter plots of predicted versus actual values and residual plots to identify any underlying patterns or areas for potential improvement. For instance, in this evaluation, the calculated RMSE value is 4.52, indicating the average deviation between the predicted and actual values.

B. Analysis of LIME Explanations for Selected Instances

A detailed analysis was conducted on LIME explanations for selected instances to assess the robustness and reliability of the interpretations. In particular, the LIME explanation for the 5th instance revealed insightful findings:

1. At the 5th Instance:

a. Intercept: 24.42
b. Local Prediction: 21.34
c. Actual Prediction: 21.48

The LIME explanation highlighted the following feature contributions:

- ✓ The feature 'RM' (average number of rooms per housing) with a threshold of <= 5.90 had a negative influence (3.65) on the predicted house price.
- ✓ The feature 'PTRATIO' (pupil teacher ratio by town) with a threshold of > 20.20 had a negative influence (1.22) on the predicted house price.
- ✓ The feature 'LSTAT' (lower status of the population) with thresholds between 6.95 and 11.86 had a positive influence (1.09) on the predicted house price.
- ✓ The feature 'NOX' (nitric oxides concentration) with thresholds between 0.45 and 0.54 had a positive influence (0.94) on the predicted house price.
- ✓ The feature 'DIS' (weighted distances to five Boston employment centres) with thresholds between 3.22 and 5.08 had a negative influence (0.13) on the predicted house price.
- ✓ The feature 'AGE' (proportion of owner occupied units built prior to 1940) with thresholds between 45.92 and 76.50 had a negative influence (0.12) on the predicted house price.

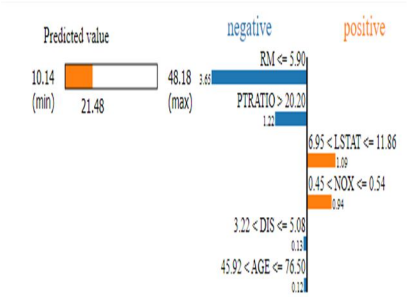


Figure3: Model_Explanation_Exp1_Sample5

d. Interpretation:

- ✓ The predicted value of the house price is 21.48 thousand dollars.
- ✓ The features 'LSTAT' and 'NOX' positively influenced the predicted house price, while 'RM', 'DIS', 'AGE', and 'PTRATIO' had negative influences.
- ✓ These findings provide valuable insights into the factors driving the model's predictions, enhancing understanding and interpretability of the predictive model.

2. At the 10th instance:

a. Intercept: 27.43
b. Local Prediction: 13.21
c. Actual Prediction: 9.78

- ✓ The LIME explanation for this instance revealed the following feature contributions:
- ✓ The feature 'LSTAT' (lower status of the population) with a threshold of > 16.46 had a negative influence (6.92) on the predicted house price.
- ✓ The feature 'RM' (average number of rooms per housing) with thresholds between 5.90 and 6.22 had a negative influence (2.98) on the predicted house price.
- ✓ The feature 'NOX' (nitric oxides concentration) with a threshold of > 0.63 had a negative influence (2.21) on the predicted house price.

- ✓ The feature 'PTRATIO' (pupil teacher ratio by town) with thresholds between 19.10 and 20.20 had a negative influence (0.91) on the predicted house price.
- ✓ The feature 'DIS' (weighted distances to five Boston employment centres) with a threshold of <= 2.11 had a negative influence (0.65) on the predicted house price.
- ✓ The feature 'AGE' (proportion of owner-occupied units built prior to 1940) with a threshold of > 93.22 had a negative influence (0.54) on the predicted house price.

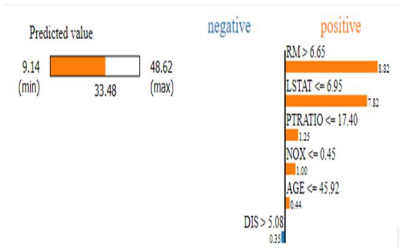


Figure4: Model_Explanation_Exp2_Sample10

d. Interpretation

- ✓ The predicted value of the house price is 9.78 thousand dollars.
- ✓ All the variables considered in the explanation have a negative influence on the predicted house prices.
- ✓ These findings provide insights into the factors contributing to the model's prediction for the 10th instance, enhancing the interpretability of the predictive model.

3. At the 20th instance

a. Intercept: 24.48
b. Local Prediction: 22.22
c. Actual Prediction: 23.75

The LIME explanation for this instance revealed the following feature contributions:

- ✓ The feature 'RM' (average number of rooms per housing) with thresholds between 6.22 and 6.65 had a negative influence (2.70) on the predicted house price.
- ✓ The feature 'PTRATIO' (pupil teacher ratio by town) with thresholds between 19.10 and 20.20 had a negative influence (1.01) on the predicted house price.
- ✓ The feature 'LSTAT' (lower status of the population) with thresholds between 6.95 and 11.86 had a positive influence (0.99) on the predicted house price.
- ✓ The feature 'NOX' (nitric oxides concentration) with thresholds between 0.45 and 0.54 had a positive influence (0.65) on the predicted house price.
- ✓ The feature 'DIS' (weighted distances to five Boston employment centres) with thresholds between 3.22 and 5.08 had a negative influence (0.23) on the predicted house price.
- ✓ The feature 'AGE' (proportion of owner occupied units built prior to 1940) with thresholds between 45.92 and 76.50 had a positive influence (0.05) on the predicted house price.

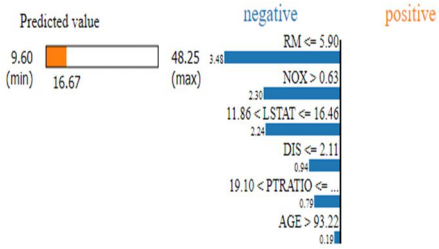


Figure5: Model_Explanation_Exp3_Sample20

- ✓ The predicted value of the house price is 23.75 thousand dollars.
- ✓ The features 'LSTAT' and 'NOX' positively influenced the predicted house price, while 'RM', 'AGE', 'PTRATIO', and 'DIS' had negative influences.
- ✓ These findings offer insights into the factors contributing to the model's prediction for the 20th instance, enhancing the interpretability of the predictive model.

4. At the 25th instance:

a. Intercept: 26.32
b. Local Prediction: 16.38
c. Actual Prediction: 16.67

The LIME explanation for this instance revealed the following feature contributions:

- ✓ The feature 'RM' (average number of rooms per housing) with a threshold of <= 5.90 had a negative influence (3.48) on the predicted house price.
- ✓ The feature 'NOX' (nitric oxides concentration) with a threshold of > 0.63 had a negative influence (2.30) on the predicted house price.
- ✓ The feature 'LSTAT' (lower status of the population) with thresholds between 11.86 and 16.46 had a negative influence (2.24) on the predicted house price.
- ✓ The feature 'DIS' (weighted distances to five Boston employment centres) with a threshold of <= 2.11 had a negative influence (0.94) on the predicted house price.
- ✓ The feature 'PTRATIO' (pupil teacher ratio by town) with thresholds between 19.10 and 20.20 had a negative influence (0.79) on the predicted house price.
- ✓ The feature 'AGE' (proportion of owner-occupied units built prior to 1940) with a threshold of > 93.22 had a negative influence (0.19) on the predicted house price.

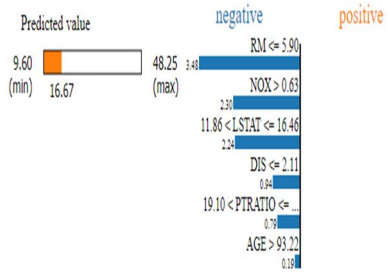


Figure6: Model_Explanation_Exp4_Sample25

d. Interpretation

- ✓ The predicted value of the house price is 16.67 thousand dollars.
- ✓ All the variables considered in the explanation have a negative influence on the predicted house prices.
- ✓ These findings contribute to understanding the factors influencing the model's prediction for the 25th instance, thereby enhancing the interpretability of the predictive model.

C. Summary of findings

In summary, this study leveraged Local Interpretable Model-agnostic Explanations (LIME) as a tool to interpret the predictions generated by a Random Forest Regressor model trained on a housing dataset. By delving into the intricate details of LIME explanations for carefully selected instances, we were able to unravel crucial insights underlying the predictive capabilities of the model. Notably, our analysis shed light on the pivotal role played by various features, including socioeconomic indicators, environmental metrics, and housing attributes, in shaping the predicted house prices. These findings underscored the multifaceted nature of the factors influencing housing market dynamics and provided valuable insights into the intricate interplay between diverse variables and their impact on property valuations.

VII Conclusion And Future Direction

A. Conclusion

In conclusion, the research utilized Local Interpretable Model-agnostic Explanations (LIME) to interpret predictions made by a Random Forest Regressor model on a housing dataset. Through detailed analysis of LIME explanations for selected instances, we gained valuable insights into the factors driving the model's predictions. Across various instances, we observed consistent patterns where certain features positively or negatively influenced the predicted house prices. For instance, features such as 'LSTAT' (lower status of the population) and 'NOX' (nitric oxides concentration) tended to have a positive influence on house prices, while features like 'RM' (average number of rooms per housing) and 'PTRATIO' (pupil-teacher ratio by town) had a negative influence. These findings underscore the importance of model interpretability in enhancing trust and transparency in

predictive models, particularly in complex domains like real estate. By employing techniques like LIME, researchers and practitioners can gain deeper insights into model predictions and make more informed decisions based on the underlying factors driving those predictions. Future research could explore the application of LIME in other domains and investigate its effectiveness in different types of predictive models. Overall, this study contributes to advancing the field of model interpretation and lays the foundation for further exploration in this area.

B. Challenges and Limitations of LIME Despite its effectiveness

LIME has certain challenges and limitations. One limitation is the reliance on proxy models, which may not fully capture the complexity of the original model. Additionally, LIME explanations may vary depending on the choice of hyperparameters and sampling techniques, introducing variability in interpretation results. Addressing these challenges requires careful consideration and further research to improve the robustness and reliability of LIME explanations.

C. Future Research Directions in Model Interpretability

Future research in model interpretability could focus on addressing the limitations of existing techniques and developing more advanced methods for explaining black box models. This includes exploring alternative explanation techniques, integrating domain knowledge into interpretation frameworks, and evaluating the impact of interpretation methods on decision-making processes. Additionally, research could investigate the application of interpretability techniques in emerging domains such as healthcare, finance, and autonomous systems, where model transparency and trust are critical. By advancing the field of model interpretability, researchers can enhance the transparency, accountability, and trustworthiness of machine learning systems in various applications.

REFERENCES

[1] Molnar, C., Casalicchio, G., Bischl, B. (2020). Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges. In: Koprinska, I., et al. ECML PKDD 2020 Workshops. ECML PKDD 2020. Communications in Computer and Information Science, vol 1323. Springer, Cham. https://doi.org/10.1007/978-3-030-65965-3_28

[2] Marchese Robinson, R. L., Palczewska, A., Palczewski, J., & Kidley, N. (2017). Comparison of the predictive performance and interpretability of random forest and linear models on benchmark data sets. Journal of chemical information and modeling, 57(8), 1773-1792.

[3] J. Zhang, Y. Wang, P. Molino, L. Li and D. S. Ebert, "Manifold: A Model-Agnostic Framework for Interpretation and Diagnosis of Machine Learning Models," in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 364-373, Jan. 2019, doi: 10.1109/TVCG.2018.2864499.

[4] Zafar MR, Khan N. Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability. Machine Learning and Knowledge Extraction. 2021; 3(3):525-541. https://doi.org/10.3390/make3030027

[5] ElShawi, R., Sherif, Y., Al-Mallah, M., Sakr, S. (2019). ILIME: Local and Global Interpretable Model-Agnostic Explainer of Black-Box Decision. In: Welzer, T., Eder, J., Podgorelec, V., Kamišalić Latifić, A. (eds) Advances in Databases and Information Systems. ADBIS 2019. Lecture Notes in Computer Science(), vol 11695. Springer, Cham. https://doi.org/10.1007/978-3-030-28730-6_4

[6] B. O. Taiwo et al., "Development of mathematically motivated artificial intelligence models for the prediction of carbonate rock lime saturation factor for cement production," Engineering Applications of Artificial Intelligence, vol. 127, part A, p. 107444, 2024, ISSN 0952-1976, doi: 10.1016/j.engappai.2023.107444.

[7] R. Saleem et al., "Explaining deep neural networks: A survey on the global interpretation methods," Neurocomputing, vol. 513, pp. 165-180, 2022, ISSN 0925-2312, doi: 10.1016/j.neucom.2022.09.129.

[8] Parisineni, S.R.A., Pal, M. Enhancing trust and interpretability of complex machine learning models using local interpretable model agnostic shap explanations. Int J Data Sci Anal (2023). https://doi.org/10.1007/s41060-023-00458-w

[9] D. Garreau and U. Luxburg, "Explaining the Explainer: A First Theoretical Analysis of LIME," in Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics, vol. 108, pp. 1287-1296, 2020. Available: https://proceedings.mlr.press/v108/garreau20a.html.

[10] Giorgio Visani, Enrico Bagli, Federico Chesani, Alessandro Poluzzi & Davide Capuzzo (2022) Statistical stability indices for LIME: Obtaining reliable explanations for machine learning models, Journal of the Operational Research Society, 73:1, 91-101, DOI: 10.1080/01605682.2020.1865846

[11] Sokol, K., Flach, P. One Explanation Does Not Fit All. Künstl Intell 34, 235–250 (2020). https://doi.org/10.1007/s13218-020-00637-y

[12] R. Knutti, R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, "Challenges in Combining Projections from Multiple Climate Models," Print Publication: 15 May 2010, pp. 2739–2758, doi: 10.1175/2009JCLI3361.1.

[13] McCreath, E., Sharma, A. (1998). LIME: A System for Learning Relations. In: Richter, M.M., Smith, C.H., Wiehagen, R., Zeugmann, T. (eds) Algorithmic Learning Theory. ALT 1998. Lecture Notes in Computer Science(), vol 1501. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-49730-7_25

[14] Chudasama, Yashrajsinh et al. ‘InterpretME: A Tool for Interpretations of Machine Learning Models over Knowledge Graphs’. 1 Jan. 2024 : 1 – 21.
[15] Prabhu, H., Sane, A., Dhadwal, R., Parlikkad, N. R., & Valadi, J. K. (2023). Interpretation of drop size predictions from a random forest model using local interpretable model-agnostic explanations (LIME) in a rotating disc contactor. Industrial & Engineering Chemistry Research, 62(45), 19019-19034.
[16] Abdullah TAA, Zahid MSM, Ali W, Hassan SU. B-LIME: An Improvement of LIME for Interpretable Deep Learning Classification of Cardiac Arrhythmia from ECG Signals. Processes. 2023; 11(2):595. https://doi.org/10.3390/pr11020595