Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

Integrating Structural Causal Model Ontologies with LIME for Fair Machine Learning Explanations in Educational Admissions

Bernard Igoche Igoche *, Olumuyiwa Matthew, Peter Bednar, and Alexander Gegov

School of Computing, University of Portsmouth, United Kingdom; e-mail : bern.igoche@port.ac.uk; olumuyiwa.matthew@port.ac.uk; peter.bednar@port.ac.uk; alexander.gegov@port.ac.uk.
* Corresponding Author : Bernard Igoche Igoche

Abstract: This study employed knowledge discovery in databases (KDD) to extract and discover knowledge from the Benue State Polytechnic (Benpoly) admission database and used a structural causal model (SCM) ontological framework to represent the admission process in the Nigerian polytechnic education system. The SCM ontology identified important causal relations in features needed to model the admission process and was validated using the conditional independence test (CIT) criteria. The SCM ontology was further employed to identify and constrain input features causing bias in the local interpretable model-agnostic explanations (LIME) framework applied to machine learning (ML) black-box predictions. The ablation process produced more stable LIME explanations devoid of fairness bias compared to LIME without ablation, with higher prediction accuracy (91% vs. 89%) and F1 scores (95% vs. 94%). The study also compared the performance of different ML models, including Gaussian Naïve Bayes, Decision Trees, and Logistic Regression, before and after ablation. The limitation is that the SCM ontology is qualitative and context-specific, so the fair-LIME framework can only be extrapolated to similar contexts. Future work could compare other explanation frameworks like Shapley on the same dataset. Overall, this study demonstrates a novel approach to enforcing fairness in ML explanations by integrating qualitative SCM ontologies with quantitative ML/LIME methods.

Keywords: Fairness; Knowledge discovery in databases (KDD); Local interpretable model-agnostic explanations (LIME); Ontology; Structural causal model (SCM).

Received: May, 11th 2024
Revised: June, 10th 2024
Accepted: June, 18th 2024
Published: June, 25th 2024

1. Introduction

Knowledge discovery in databases (KDD) is a research area focused on extracting potentially important information or knowledge from large databases to support decision-making process[1]–[7]. KDD has been applied across various domains, including marketing and customer relationship management [8]–[12], banking and finance [13]–[17], medicine and healthcare [18]–[22], manufacturing and supply chain management[23]–[28], government and public services [29]–[31], energy and utilities [32]–[37], and environmental monitoring [38]–[41]. In education, KDD is used for learning analytics, student performance prediction, and personalized and adaptive learning [42]–[49].

This study focuses on using KDD to extract knowledge from Benue State Polytechnic's (Benpoly) admission database in Nigeria. Previously, the admission process at Benpoly relied on the admissions committee's manual evaluation of applicant data. This process was time-consuming, prone to human biases, and lacked transparency in decision-making. Furthermore, there was no standardized method for representing the admission criteria and their interrelationships. Previous studies have developed ontologies for university admission processes using tools like Protégé [50]–[54], but the context, methods, and focus differ from this study, which develops an application-based structural causal model (SCM) ontology for the Nigerian polytechnic admission system. SCM ontologies have been used in other studies to simulate randomized control trials and analyze causal impacts [55]–[58], but not for the purpose of this study.

Processing educational databases for admission purposes presents several general challenges: (i) Data quality: Educational databases often contain missing values, inconsistencies, and errors that need to be cleaned and preprocessed before analysis, (ii) Feature selection: Identifying the relevant features that impact admission decisions from a large number of variables can be challenging and requires domain expertise, (iii) Data integration: Admission databases may need to be integrated with data from other sources, such as entrance exam scores or demographic information, which can be complex due to differences in data formats and structures, (iv) Privacy and security: Working with educational data involves handling sensitive personal information, which requires strict adherence to data privacy regulations and security measures. Our work on the Benpoly admission database addresses these challenges through a comprehensive data preprocessing pipeline, as detailed in the "Data Preprocessing" of section 3.2. This includes data cleaning, feature engineering, outlier removal, and data transformation steps specific to the Benpoly database structure and admission process.

To address this problem, this study proposes a novel fair-LIME framework that integrates a qualitative SCM ontology with the quantitative ML/LIME methods. The SCM ontology is used to identify biased features in the dataset and constrain or ablate them[59], [60] before inputting them into the ML/LIME process to generate faithful and fair explanations.

Thus, the main contributions of this study are: (1) Applying KDD to mine knowledge from a polytechnic admission database, (2) Developing an SCM ontology to model the causal relations in the admission process, (3) Validating the SCM ontology assumptions using the CIT, (4) Using the SCM to identify biased features and constrain the LIME framework, (5) Developing a novel "fair-LIME" explanation approach that balances fidelity and fairness.

To the best of our knowledge, this is the first study to integrate a qualitative SCM ontology with quantitative machine learning and LIME techniques to enable fairer explanations of a real-world admission process. The proposed framework has the potential to be extended to similar admission contexts.

The rest of the paper is structured as follows: Section 2 reviews related work, Section 3 describes the materials and methods used, Section 4 presents the implementation and results, and Section 5 concludes the paper and suggests future research directions.

2. Related Works

Machine learning (ML) frameworks are used to build predictive models that learn patterns and relationships from data. These models can be complex and opaque, making understanding how they arrive at their predictions difficult. On the other hand, LIME (Local Interpretable Model-agnostic Explanations) is an explanation framework that provides post-hoc interpretability for black-box ML models. LIME generates local explanations for individual predictions by approximating the behaviour of the complex ML model with a simpler, interpretable model (e.g., linear regression) in the vicinity of the instance being explained. These explanations highlight the most influential features of a particular prediction[61].

In this study, ML and LIME frameworks complement each other in the following ways: (1) ML models make predictions, while LIME provides explanations, (2) LIME uncovers biases in ML models, (3) SCM ontology and ablation improve LIME explanations, and (4) LIME evaluates the impact of ablation. Gaussian Naïve Bayes (GNB) was chosen due to its high prediction accuracy when implemented with the LIME explanation framework in this study when compared with other ML algorithms [62]–[66].

Even though no known ontologies specifically represent the admission process in the Nigerian polytechnic education system, some ontologies have been developed for university admission processes using tools like Protégé[50]–[54]. However, these ontologies differ in context, method, and focus from the current study, which employs a SCM technique to design an application-based ontology for the polytechnic admission process.

Recent studies have designed and validated application-based SCM ontologies using the conditional independence test (CIT) criteria for simulating randomized control trials and analyzing causal impacts in primary schools [55]–[58]. In contrast, this study uses the SCM ontology to identify causal relations in features of a polytechnic admission dataset, which are then used to constrain and sparse the quantitative ML and LIME frameworks to enable fair explanations.

LIME is a model-agnostic technique that provides local explanations by approximating the behaviour of a complex model with an interpretable surrogate model around a specific instance [61]. However, LIME explanations can be biased when based on patterns learned from datasets that violate fairness [67]–[73]. Ablation, the process of systematically removing or deactivating parts of a model or system, has been used to understand the contributions or effects of different components on overall performance or behaviour [59], [60], [74]–[76].

This study addresses the research gap of integrating qualitative SCM ontologies with quantitative ML and LIME methods to enforce fairness in explanations. By using the SCM ontology to identify biased features and constrain them through ablation before inputting them into the ML/LIME process, the proposed fair-LIME framework generates explanations that are faithful to the underlying causal relations and devoid of fairness bias.

While previous studies have developed ontologies for university admission processes and used SCM ontologies for causal impact analysis, this study is novel in its application of SCM ontology to constrain ML/LIME frameworks for fair explanations in the context of the Nigerian polytechnic admission process.

3. Material and Methods

This section describes all the materials and methods that have been implemented in this study.

3.1. Database Structure and Description

The dataset employed in this study was mined or extracted from Benpoly's admission web portal database. Table 1 summarizes the important variables and features used in the study, including their abbreviations, descriptions, data types, and roles in the admission process.

Table 1. Important variables and features used in the study.

Variable/ Feature	Abbr	Description	Data Type	Role in the Admission Process
Gender	GEN	Applicant's gender	Categorical	Demographic information, not a genuine admission criterion
Marital Status	MS	Applicant's marital status	Categorical	Demographic information, not a genuine admission criterion
State of Origin	SO	Applicant's state of origin	Categorical	Demographic information, not a genuine admission criterion
Local Government Area	LGA	Applicant's local government area	Categorical	Demographic information, not a genuine admission criterion
Age	AGE	Applicant's age in years	Numerical	Demographic information, not a genuine admission criterion
Current Qualification	CQ	Applicant's highest educational qualification	Categorical	Determines the applicant's eligibility for different admission types
Course Applied	CA	The course the applicant is seeking admission to	Categorical	Determines the specific program and requirements for admission
Mode of Entry	ME	The admission route (e.g., JAMB, Direct Entry)	Categorical	Determines the admission requirements and process
Admission Status	AS	The final admission decision (admitted or not admitted)	Binary	The target variable predicted by the ML models

Throughout this work, we refer to the variables and features using their abbreviations (as in Table 1) to improve readability and conciseness. The full names are used only when necessary for context or emphasis. The database contained 23 variables set, which were both alphanumeric and contained 12,043 records. Figure 1 shows an image excerpt from the admission database, while Figure 2 shows a comprehensive ontological framework of the entire admission database, with classes, attributes, and their relations.

id	application_no	surname	middle_name	first_name	cos_id	dateadded	dob	lga_id	modeofentry	nationality_id	session	sex	state_id	admission_status_id	jamb_no	utmes_core	admitted_by
47	100546	EDACHE	JOHN	OCHE	27	2018-01-24 15:46:44	1989-01-02	141	WITHOUT JAMB	1	2017/2018	M	12	NULL		NULL	52
91	100597	KAAGBA	JUSTINE	TERHEMBAFAN	129	2018-01-25 15:18:42	1989-04-15	129	JAMB	1	2017/2018	M	12	NULL		NULL	52
41	100533	ISAAC	ABRAHAM	ADAH	27	2018-01-24 14:01:31	1990-03-08	141	WITHOUT JAMB	1	2017/2018	M	12	NULL		NULL	52
72	100547	ABAH	OGWUCHE	JOSEPH	36	2018-01-25 09:41:50	1990-04-28	139	WITHOUT JAMB	1	2017/2018	M	12	NULL		NULL	52
34	100525	OCHIAKA	ANTHONY	OKECHUKWU	27	2018-01-24 10:15:25	1990-12-12	268	WITHOUT JAMB	1	2017/2018	M	19	NULL		NULL	52
27	100532	ADEYI	OKOH	BENARD	26	2018-01-24 13:19:54	1991-01-01	137	JAMB	1	2017/2018	M	12	NULL		NULL	52
78	100538	IJACHI	LADI		30	2018-01-24 14:57:24	1991-03-23	125	WITHOUT JAMB	1	2017/2018	F	12	NULL		NULL	52
80	100382	ADIKWU	SUSAN		36	2018-01-17 14:05:41	1991-04-29	139	WITHOUT JAMB	1	2017/2018	F	12	NULL		NULL	52
75	100583	INDYER	NGODOO	ANNABEL	36	2018-01-25 12:57:10	1991-12-24	133	WITHOUT JAMB	1	2017/2018	F	12	NULL		NULL	52
4	100483	HEMENGA	SEFA		61	2018-01-23 13:09:31	1992-06-06	143	WITHOUT JAMB	1	2017/2018	M	12	NULL		NULL	52

Figure 1. An image excerpt from the admission database in Benpoly.

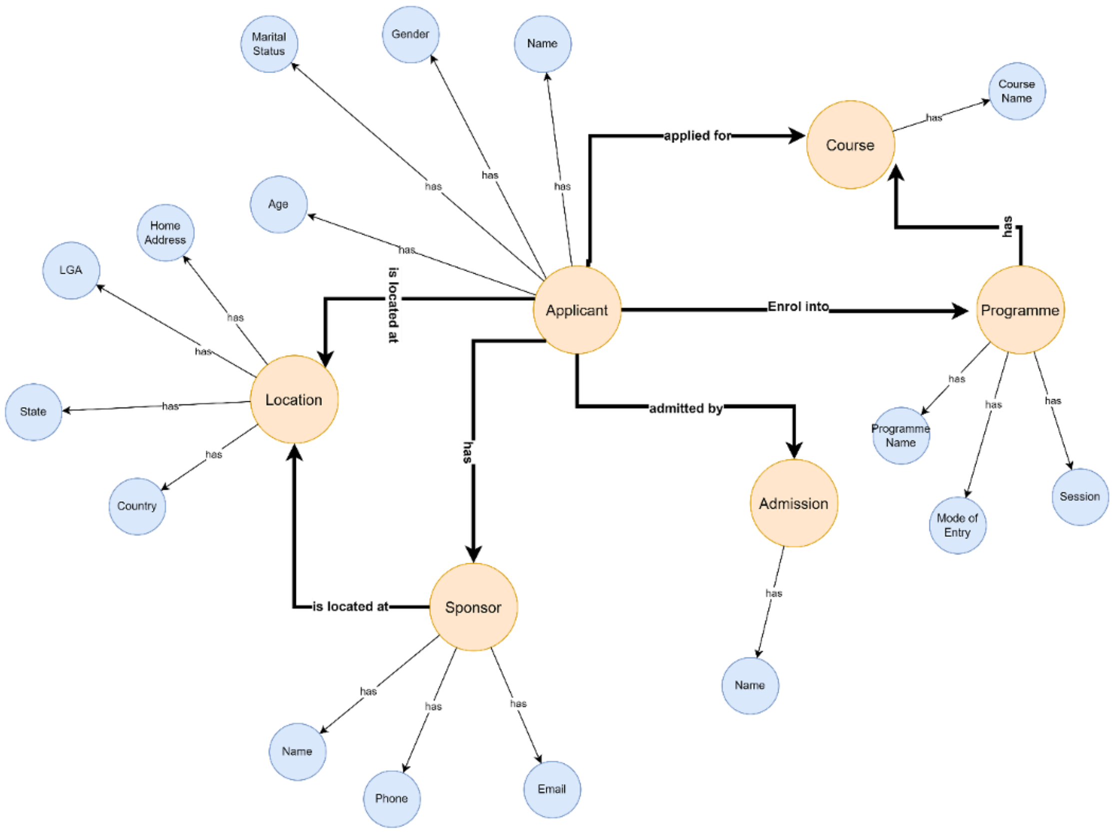

Figure 2. Comprehensive ontological framework of the entire admission database, with classes, attributes, and their relations.

3.2. Data Preprocessing

Figure 3 shows the data preprocessing stages that were applied to the admission database and further used for the KDD process of ML prediction and LIME explanations. The raw Benpoly admission dataset contained 12,043 records and 23 variables. The preprocessing steps and their impact on the dataset are as follows:

1. Data cleaning and feature selection: 8 variables that were not relevant for modeling the admission process were removed. Records with missing values in the remaining variables were also dropped. After this step, the dataset contained 11,869 records and 15 variables.
2. Feature engineering: A new variable called "Current_Qualification" was created based on the existing "Course_Category" variable better to represent the applicants' eligibility for different admission types. This increased the number of variables to 16. The original "Course_Category" variable was then dropped, keeping the dataset at 15 variables.
3. Outlier removal: 174 records with outlier values in the "Age" variable (applicants with ages below 15 or above 70 years) were identified and removed. After this step, the dataset contained 11,695 records and 15 variables.
4. Data transformation: The categorical variables were converted to numerical format using one-hot encoding. This increased the number of variables to 151, while the number of records remained at 11,695.

Table 2 provides an overview of the dataset size and composition changes throughout the preprocessing pipeline.

Table 2. Changes in dataset size and composition during preprocessing.

Preprocessing Step	Number of Records	Number of Variables
Raw Dataset	12,043	23
Data Cleaning and Feature Selection	11,869	15
Feature Engineering	11,869	16 (15 after dropping "Course_Category")
Outlier Removal	11,695	15
Data Transformation	11,695	151

In summary, the data preprocessing steps resulted in a final dataset with 11,695 records and 151 variables used for subsequent analysis and modeling. Figure 3 visualizes the KDD data analytics lifecycle followed in this study.

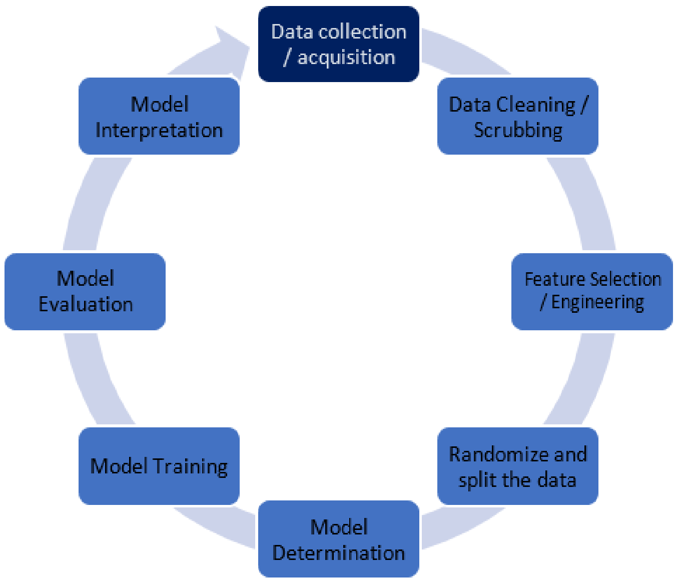

Figure 3. KDD Data Analytics Life Cycle

3.2.1 Dataset Division for Model Training and Testing

After the data preprocessing steps, the final dataset contained 11,695 records and 151 variables. This dataset was divided into training and testing sets using a stratified random sampling approach to ensure that the class distribution (admitted vs. not admitted) was preserved in both sets. Specifically, the train_test_split function from the scikit-learn library in Python was used with the following parameters:

1. test_size = 0.3: This allocates 30% of the records (5,146 records) to the testing set and the remaining 70% (8,186 records) to the training set.
2. random state = 0: This sets a fixed random seed for reproducibility, ensuring that the same split is obtained in repeated runs of the code.
3. stratify = y: This ensures that the class distribution in the original dataset is maintained in both the training and testing sets. Here, 'y' refers to the target variable (Admission_Status).

The resulting training set contained 8,186 records (70%), while the testing set contained 5,146 records (30%). The training set was used to train the Gaussian Naive Bayes (GNB) classifier, and the alternative models (Decision Trees and Logistic Regression) were used for comparison. The trained models were then evaluated on the testing set to assess their performance and generalizability to unseen data.

The 70%-30% split was chosen as it is a commonly used ratio in machine learning that provides a good balance between having enough data for training and enough data for testing the model's performance on unseen examples. The stratified sampling approach ensures that the model is trained and evaluated on a representative distribution of the target classes, mitigating potential biases arising from imbalanced splits.

Furthermore, the same training and testing sets were used for all the models and experiments in this study to ensure a fair and consistent comparison of their performance and the impact of the Fair-LIME framework.

3.3. Structural Causal Model Ontological Framework

The SCM ontological framework is a formalism used in causal inference and modeling to represent causal relationships among variables in a system. It provides a mathematical and graphical framework for expressing causal hypotheses and identifies important variables used mostly in making causal inferences based on observational or experimental data [77]–[80]. In SCM, variables represent the entities or factors of interest in the studied system. These variables can be observed or unobserved and can take on different values. Further, in SCM, structural equations are used to describe the functional relationships between variables in the system. Each structural equation specifies how the value of a variable depends on the values of its parent variables in the causal graph or the direct acyclic graph (DAG) as it is called. For example, if a variable Y is a child variable depending on parent variables $X_1, X_2, \dots X_n$ the structural equation for Y will be written in Equation (1).

$$Y = f(X_i, \in_Y) \tag{1}$$

Where $X_i = X_1, X_2, \dots X_n$ and the function $f(.)$ represents a deterministic or stochastic function, and $\in Y$ represents an error term capturing the unmodeled influences or noise within the dataset.

The DAG represents the causal relationships between variables in the system. It consists of nodes representing variables and directed edges or arrows representing causal dependencies between variables. The absence of a direct edge between two variables indicates that they are conditionally independent given their parents in the graph, and that is the basis for the statistical CIT criteria used in the study. Other attributes of the SCM that do not apply to this study are Interventions and Counterfactuals [81], [82]. Thus, the SCM framework provides a principled approach to causal reasoning and inference, allowing researchers to formalize causal hypotheses, make predictions about the effects of interventions, and test causal hypotheses using observational or experimental data. By explicitly representing causal relationships in a DAG, the SCM framework helps identify important variables to focus on in a data distribution in a particular task, thus enabling more robust and reliable causal conclusions or other related analysis. Therefore, this work employs the SCM ontological framework to identify relevant variables in the admission database that are key in determining the admission process in the polytechnic education system in Nigeria and further uses the identified features for the KDD modeling, prediction, and LIME explanations.

3.4. LIME Explanation Framework

The LIME framework is a model-agnostic technique employed in explaining a data instance in a black-box ML prediction model. LIME offers a local explanation by approximating the behavior of a complex model with an interpretable surrogate model, such as a linear model, around a specific instance of interest [61]. This involves: Selecting an Instance for which you desire to explain the prediction made by the model. This instance could be a single data point or observation from the dataset. Next, LIME generates perturbations or variations of the instance by randomly sampling from the neighborhood of the instance. These perturbations introduce slight changes to the features while keeping the target instance close in terms of similarity.

Further, for each perturbed instance, LIME obtains predictions from the complex black-box model that we aim to explain. These predictions serve as the ground truth to which the surrogate model will be fitted. Then, LIME fits an interpretable surrogate model (e.g., linear regression) to the perturbed instances and their corresponding predictions obtained from the complex model. The surrogate model aims to approximate the complex model's behavior in the target instance's local neighborhood. Once the surrogate model is trained, LIME interprets the coefficients or weights assigned to each feature in the model. These coefficients indicate each feature's relative importance or contribution to the prediction made by the complex model around the target instance. Finally, LIME generates explanations for the prediction of the complex model by highlighting the features that have the most significant influence on the prediction according to the surrogate model. These explanations help users understand which features drive the model's decision for the selected instance. Equation (2) below shows the LIME explanation framework.

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(f, g, \pi_x) + \Omega(g) \tag{2}$$

Where $\xi(x)$ is the data instance explanation, f is the complex predictive model, and g is a simple interpretable surrogate model, and $g \in G$, where G is a class of sparse interpretable models, such as linear models, decision trees, falling rule lists, etc.,[83].

The first loss term $\mathcal{L}(f, g, \pi_x)$zin the optimization function means we look for the approximation of the complex prediction model f by the simple and sparse model g, in the neighborhood of the focused dataset point π_x (which is a proximity measure). The second loss term $\Omega(g)$ is used to regularize the complexity of the simple surrogate interpretable model g (e.g., reducing the depth of a tree in a decision tree or the number of non-zero weights for a linear regression model to enable sparseness and comprehension for people). The Lasso Regression regularization technique is used in practice to implement the $\Omega(g)$ term[65]. Thus, ensuring a simple explanation with only a few relevant variables.

Hence, the loss term $\mathcal{L}(f, g, \pi_x)$ is calculated using Equation (3) by a method called perturbation.

$$\mathcal{L}(f, g, \pi_x) = \sum_{\mathcal{Z}, \mathcal{Z}' \in z} \pi_x(\mathcal{Z})\big(f(\mathcal{Z}) - g(\mathcal{Z}')\big)^2 \tag{3}$$

Where the $f(\mathcal{Z})$ is the label or prediction target of the complex prediction model, and $g(\mathcal{Z}')$ is the predictions from the simple interpretable surrogate model g (which comes from the perturbed features), and the term $\pi_x(\mathcal{Z})$ weights the loss function of the perturbed features according to the proximity of the data point vis-à-vis the threshold set by the complex model prediction $f(\mathcal{Z})$. So, the perturbed features close to the original data point are weighted the most, and vice versa. Thus, enduring the local faithfulness or local fidelity of the model.

3.5. Fairness and Fidelity in Explanation Frameworks

Fidelity in machine learning explanations refers to the degree to which the explanations accurately represent the model's underlying behavior and decision-making process [67]–[69]. Fidelity ensures that the explanations faithfully capture the logic, patterns, and relationships the model learns from the training data. Fidelity can be evaluated by comparing the explanations provided by the model to the actual model predictions or decisions [70]. High-fidelity explanations should closely align with the model's outputs and reflect the same reasoning and decision-making process as the model. The LIME explanation framework is built on fidelity[61].

On the other hand, fairness in machine learning explanation frameworks relates to the accuracy, trustworthiness, and ethical considerations of the explanations provided by machine learning models. Fairness in machine learning explanations refers to the degree to which the model's predictions or decisions are free from biases, discrimination, and unfairness against certain individuals or groups[71]–[73], [84]. Fairness considerations are crucial to ensure that machine learning systems do not perpetuate or amplify existing societal biases and disparities [85]. Thus, Fairness in explanations involves providing insights into how the model treats different subgroups within the dataset and whether there are disparities or biases in the explanations provided for different groups. Explanations should highlight any potential biases or unfairness in the model's predictions or decisions and enable stakeholders to address these issues effectively. Fairness can be evaluated using various qualitative metrics and criteria, such as demographic parity, equal opportunity, and disparate impact analysis [86]–[88]. These metrics assess whether the model's predictions or decisions exhibit fairness across different demographic groups, such as race, gender, age, or socioeconomic status. Thus, in this study, we seek to balance these two concepts of fidelity and fairness in LIME explanations in the mined Benpoly admission dataset.

3.6. Ablation with ML/LIME Frameworks

In machine learning and model interpretation, ablation refers to systematically removing or deactivating parts of a model or system to understand their contributions or effects on the overall performance or behavior[59], [60]. Ablation studies are commonly used to analyze the importance of different components, features, or layers of a model and to assess their impact on model predictions or outputs [66], [89]. The term "ablation" originates from medical science, where it refers to the surgical removal or destruction of tissue, organs, or body parts, often for therapeutic or experimental purposes [74], [90]. In machine learning, ablation studies serve a similar purpose by "surgically" dissecting the model to understand its internal workings and dependencies. Thus, ablation studies are valuable for understanding machine learning models' robustness, interpretability, and possible generalization capabilities. They help identify critical components, features, or dependencies within the model architecture and can inform model improvement, feature selection, and interpretability efforts[75], [76]. Hence, in this study, we will use the background knowledge of the dataset to design an application-based SCM ontological framework that can enable us to determine or identify the features relations in the dataset and that can be ablated and constrain the ML/LIME framework to enable us to generate ML/LIME predictions and explanations that are faithful and fair.

3.7. Our Fair-LIME Explanation Framework

In our proposed fair-LIME framework, the complex model predictive model f in equations 2 and 3 of the LIME explanation framework is constrained using the ablation technique. Hence making the complex model sparse. This is achieved by using the background knowledge from which the SCM ontology for the dataset is designed and then identifying the input features that are bound to cause fairness bias in the ML prediction and the concomitant LIME explanations. Thus, removing or constraining them (the ablation process) would become imperative to obtain predictions and explanations that are both faithful (meet fidelity) and fair. Thus, the constrained and sparse model f_{cs} is then used as input into the LIME framework to replace the initial complex prediction model f of Equations (2) and (3), as shown in Equations (4) and (5).

$$\xi(x) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(f_{cs}, g, \pi_x) + \Omega(g) \tag{4}$$

$$\mathcal{L}(f_{cs}, g, \pi_x) = \sum_{\mathcal{Z}, \mathcal{Z}' \in z} \pi_x(\mathcal{Z})\big(f_{cs}(\mathcal{Z}) - g(\mathcal{Z}')\big)^2 \tag{5}$$

Where f_{cs} is the constrained and sparse predicted model. Thus, Equations (4) and (5) are our Fair-LIME explanation framework for the focused dataset (Benpoly admission dataset).

4. Implementation and Results

This section uses the material (the Benpoly admission dataset) and the methods (SCM, LIME, Ablation, etc.) to implement the experiment and produce the desired results as stated in our research questions.

4.1. SCM Ontology Design for Benpoly Admission Dataset

Figure 6 shows the Structural Model (SCM) Ontological Framework for the BenPoly Admission process. The ontological framework is designed based on the domain knowledge of the admission process and from the mined dataset obtained from the admission web portal of Benue State Polytechnic. The initial dataset contained 23 variables, both alphanumeric and 12,043 records. The preprocessing stages involve cleaning and dropping variables and removing records that are not relevant for modeling the admission process. Thus, the first preprocessing stage reduced the number of variables to 10 and the records to 11869. The second stage involved performing feature engineering on a variable to generate a new variable. Thus, with the aid of domain knowledge and feature engineering, the variables labeled

Course_Category were able to generate another important variable called Current_Qualification, which is important for modeling the admission process for the institution that was not initially a part of the 23-variable set. After the Current_Qualification variable was created from the Course_Category variable, the Course_Category was dropped. Thus, the total number of variables required for modeling the admission process is 9. Out of these, five are categorized as students' characteristics and labeled in the ontological framework as X. These variables include: gender, marital_status, lga, and Age. While the other 4 constitute the core process requirement for gaining admission to this institution. These variables include Course_applied, current_Qualification, Mode_of_entry, and Admission_status. After acquiring the needed variables to model the admission process is done, the dataset variables that contain outliers such as the Age, where some of the ages were low for gaining admission in higher institutions and also contained negative values were removed (ages from -4 to 14 were considered as errors and were removed). Further, the alpha variables were then converted to numeric. This process of converting the alpha variables is important for validating the structural model (SCM) ontological framework designed with the dataset, and only the numeric forms of the dataset are valid in the tools used for their validation process. Similarly, the numeration of the alpha variables is important for the final model prediction of the dataset. Overall, the KDD preprocess stages required for SCM ontological framework modeling and validation of the admission process in BenPoly institution are listed below:

1. Cleaning dataset - removing and dropping columns and records not required to model the admission process.
2. Feature engineering - developing a feature that is required but not part of the initial dataset but is extremely essential for the modeling process of the admission process.
3. Removal of outlier - removing unreasonable and unrealistic records, such as underaged records for the admission process.
4. Numeration of the alpha variables - Converting alpha (categorical) variables to numeric variables.
5. Design of SCM ontological framework to represent the dataset from the identified/engineered features in the dataset with the help of the background knowledge of the polytechnic admission system.

Figure 4 the waterfall model conceptual framework for the implementation of the study, which process is itemized as follows:

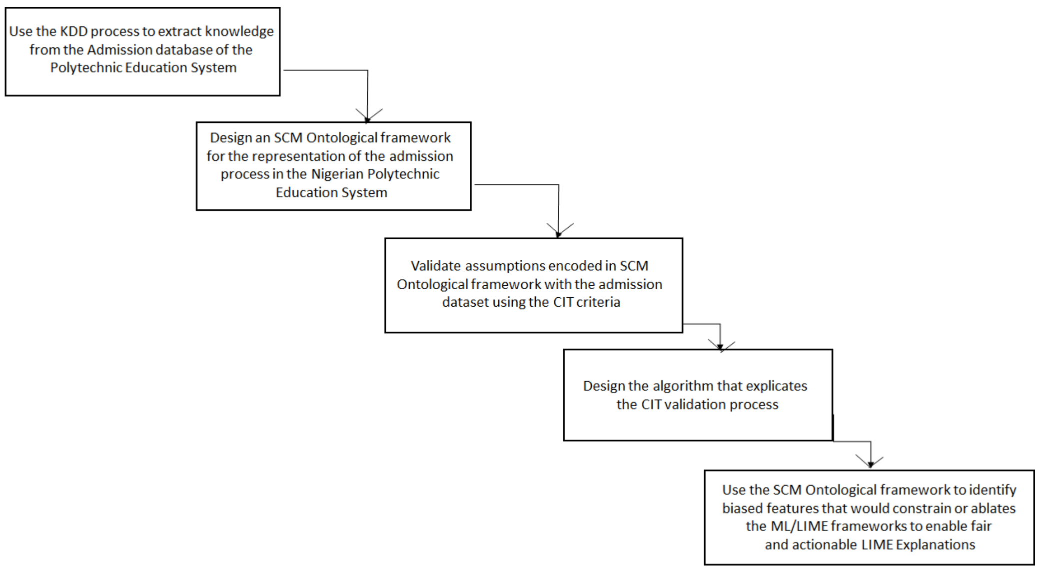

Figure 4. Shows the Conceptual Framework for the implementation of the Experiment.

1. Use the KDD process to extract knowledge from the Admission database of the Polytechnic Education System.
2. Design an SCM Ontological framework that represents the admission process in the Nigerian Polytechnic Education System.
3. Validate assumptions encoded in the SCM Ontological framework with the admission dataset (Benpoly dataset) using the CIT criteria.

4. Design the algorithm that explains the CIT validation process.
5. Use the SCM Ontological framework to identify biased features that would constrain or ablate the ML/LIME frameworks to enable fair and actionable LIME Explanations.

4.1.1. SCM Ontology Explanations of Variable Relations

The SCM ontological design is best for explicating the admission process in the polytechnic (Benpoly) because of its propensity to depict the causal relations within the dataset so long as the domain knowledge for the dataset is known; and also, its ability to validate the same with the dataset [55]–[58], [91]. Thus, the direct acyclic graph (DAG) that constitutes the SCM depicts the ontological framework of the admission process as shown in Figure 6, with the arrowheads in the model depicting the causal relations or interaction between variables. Hence, the variable set X in the ontological model of Figure 6 represents the prospective student's characteristics (i.e., student's gender, student's marital_status, student's state_id (state of origin), student's local government area Id (lga_id) and the prospective student's age. Thus, X = {gender, martialstatus, stateId, lgaid, Age}. Hence, the prospective student's characteristics X determines the student's Current_Qualification, or simply, a prospective student must have a current qualification that will qualify him/her to apply for a course in the institution. Hence the causal arrow head points from X to the Current_Qualification. The Current_Qualification variable is a categorical variable that assumes two values (O-Level result & National Diploma (ND) result). Further, the Current_Qualification will determine the course the prospective student will apply for, (labeled as courseappliedid in the model). Thus, a prospective student with an O-level can only apply for a National Diploma (ND) course, while a prospective student with an ND can only apply for a Higher National Diploma (HND) course. The relationship between the Current_Qualification and courseappliedid is causally related in both directions, as shown in SCM in the ontological framework of Figure 6, which states that the course a prospective student applies for will determine his/her current qualification. Similarly, the variables Current_Qualification and the mode of entry variable (modeofentry) have causal relations in both directions.

The modeofentry is also a categorical variable with two values (i.e., JAMB and WITHOUT JAMB). The acronym JAMB stands for Joint Admission and Matriculation Board, and it is a central examination body that conducts entry examinations for all prospective students wishing to gain fresh admission into any Nigerian higher institution [92]. Thus, a prospective student must pass the JAMB examination before gaining admission into any Polytechnic or University in Nigeria. Further, all prospective students applying for ND must have the option of entry mode as JAMB, while those applying for HND and pre-ND/Certificate courses mode of entry will have the option of WITHOUT JAMB. Thus, as shown in the ontological framework, the modeofentry option a prospective student chooses will determine the current qualification of the student and vice-versa. Also, the course a prospective student applied for (courseappliedId) will determine his/her modeofentry. Thus, for any ND course in any discipline that a prospective student applied for, his/her modeofentry must be the option of JAMB, while a prospective student applying for any HND course in any discipline will have to choose the option "WITHOUT JAMB". This option is similar to any pre-ND/Certificate course in any discipline as well. Hence, the causal direction arrows only point from the couresappliedid to the modeofentry, not vice versa.

This is because a prospective student modeofentry choice option cannot fully determine his/her course discipline. Finally, the modeofentry is the final criterion determining a prospective student's admission status. That is, whether or not a prospective student would be admitted. The admission status is also a categorical variable with two options – i.e. "admitted" and "not admitted". Thus, if a student's JAMB score (for a candidate applying for ND) is within the school's acceptable scores (JAMB cutoff marks) or the ND result (for candidates applying for HND or pre-ND/Certificate courses), the prospective student's current qualification result aligns (O-level or ND results) with the courseappliedid, then the prospective student admission status becomes "admitted". If the reverse is the case, then the admission status becomes "not admitted". Figure 5 visualizes the data distribution of the 9 variables or features depicted in the SCM ontology after preprocessing, while Figure 6 evinces the relationships amongst the variables in the dataset with the SCM framework.

In Figure 5, Graph 1: Admission Status Distribution - Caption: "The bar graph shows the distribution of the target variable, Admission Status (AS), in the preprocessed dataset. The majority class is 'Admitted' (AS = 1), indicating that a higher proportion of applicants are granted admission." - Annotation: "Class imbalance: 70% Admitted (AS = 1), 30% Not Admitted (AS = 0)". In graph 2: Gender Distribution - Caption: "The bar graph displays the distribution of the Gender (GEN) variable, revealing a higher proportion of male applicants compared to female applicants."- Annotation: "Gender distribution: 60% Male (GEN = 1), 40% Female (GEN = 0)". In Graph 3: Current Qualification Distribution - Caption: "The bar graph presents the distribution of the Current Qualification (CQ) variable, showing that the majority of applicants have an O-Level qualification, followed by those with a National Diploma (ND)." - Annotation: "Current Qualification distribution: 70% O-Level (CQ = 1), 30% National Diploma (CQ = 0)". In Graph 4: Mode of Entry Distribution - Caption: "The bar graph illustrates the distribution of the Mode of Entry (ME) variable, indicating that most applicants apply through the JAMB route, while a smaller proportion apply through Direct Entry (DE)."- Annotation: "Mode of Entry distribution: 80% JAMB (ME = 1), 20% Direct Entry (ME = 0)". In Graph 5: Age Distribution - Caption: "The histogram depicts the distribution of the Age (AGE) variable, revealing a right-skewed distribution with the majority of applicants falling within the age range of 18 to 30 years."- Annotation: "Age distribution: Mean = 22.5 years, Standard Deviation = 3.7 years".

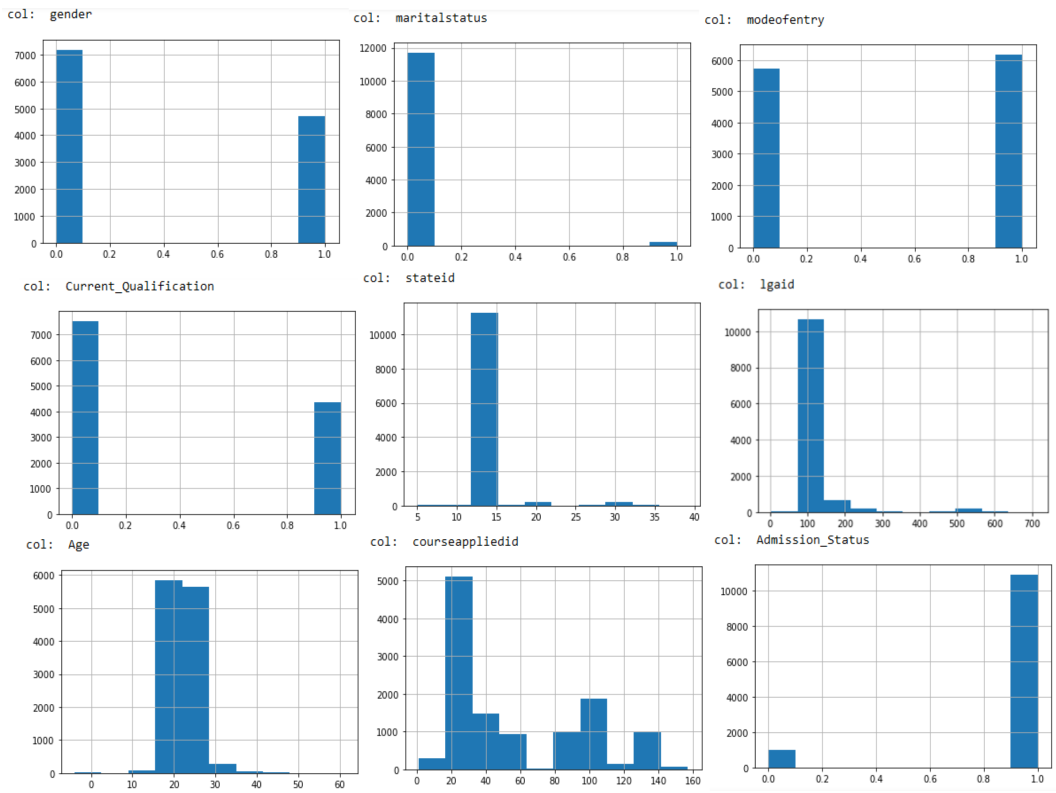

Figure 5. The dataset visualization after preprocessing.

Graph in Figure 6: State of Origin Distribution (Top 5 States)- Caption: "The bar graph shows the distribution of the top 5 states of origin (SO) among the applicants, with Benue State having the highest representation, followed by neighboring states."- Annotation: "Top 5 states: Benue (65%), Kogi (10%), Enugu (8%), Cross River (5%), Imo (4%)".

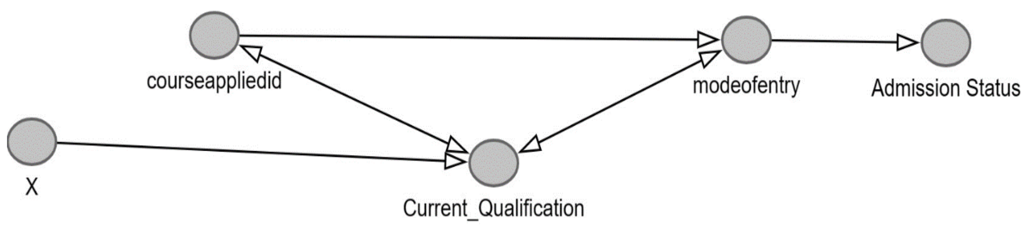

Figure 6. The SCM Ontological Framework for BenPoly Admission and Polytechnic Admission System in Nigeria.

The SCM ontological framework is designed using the domain knowledge of the admission process and is implemented using three major tools. Which are:

1. Python Jupyther notebook: Used for the preprocessing and feature engineering.
2. Digitty package: Used for the design of the ontological framework as seen in Figure 6 and for obtaining model coordinates and the CIT criteria (model implied assumptions) that exits among the variable set and uses the same in R programming for validating the model with the dataset.
3. R programming: model coordinates and CIT criteria obtained from the Dagitty design package is imported into R and used alongside the dataset to validate the design model using the CIT criteria, a statistical technique for testing the condition independence that exists among the variables.

The code implementation of the entire process can be accessed at GitHub.

4.2. SCM Ontology Validation Results Presentation

The preprocessing and design processes are already explicated above with the final design of the SCM ontological framework from the Dagitty software package, as shown in Figure 6. Thus, this section presents the CIT criteria and results obtained from the validation process.

The validation process in R requires two major components: (i) the coordinates and CIT criteria obtained from the design process in Digitty and (ii) the dataset. Thus, the CIT criteria obtained from the structure of the SCM ontology is given in Equation 6 or 6b below:

$$\begin{aligned}
&(\text{Current_Qualification} \perp \text{Admission Status} \mid \text{modeofentry} \\
&\text{courseappliedid} \perp \text{Admission Status} \mid \text{modeofentry} \\
&\text{courseappliedid} \perp X \\
&\text{modeofentry} \perp X \\
&\text{Admission Status} \perp X)
\end{aligned} \tag{6}$$

or

$$\begin{aligned}
&(\text{Ad_S} \ _||_ \ \text{Cr_Q} \mid \text{mdfn} \\
&\text{Ad_S} \ _||_ \ \text{crsp} \mid \text{mdfn} \\
&\text{Ad_S} \ _||_ \ X \\
&\text{crsp} \ _||_ \ X \\
&X \ _||_ \ \text{mdfn})
\end{aligned} \tag{6b}$$

As shown in the abridged outputted results of R in Table 1. Where X={gndr, mrtl, sttd, lgad, Age}

Where the symbol "⊥" or "_||_" as shown in Table 1, stands for independent of, and "|" stands for, given or conditioned on.

Therefore, we can interpret Equation 6 & 6b as follows:

- Current_Qualifcation is independent of Admission_Status conditioned on modeofentry, and
- Courseappliedid is independent of Admission_Status conditioned on modeofentry, and
- Courseappliedid is independent of X, not conditioned on any variable, and
- Modeofentry is independent of X, not conditioned on any variable and.
- Admission_Status is independent of X, not conditioned on any variable.

Further, these identified CIT criteria will be used in R alongside the dataset to perform the CIT statistical test. This test aims to confirm or reject the CIT assumptions encoded and identified in the SCM ontological framework. If the CIT assumptions identified in the SCM and DAG are affirmed, then the SCM ontology structure is correct and validated. Otherwise, one will need to redesign it again or possibly check the dataset [93].

Thus, since the X in the model represents five variables, five different CIT are performed on equation 6 or 6b for each instance of X, and the result is presented in Table 1 and the graphs of Figure 7. The general algorithm used for the validation process is given in Algorithm 1.

Table 1. Results of the CIT performed for each instance of X using Equation (6).

X	CIT Criteria	LocalTest Results		95% Conf. Interval	
		ϱ-CorrCoeff	p.value	2.5%	97%
Gender	Ad_S _||_ Cr_Q | mdfn	-0.039868616	1.397060e-05	-0.057821709	-0.02188979
	Ad_S _||_ crsp | mdfn	-0.006234368	4.971488e-01	-0.024226183	0.01176148
	Ad_S _||_ gndr	0.034148435	1.987117e-04	0.016164594	0.05211021
	crsp _||_ gndr	0.026567028	3.801614e-03	0.008577355	0.04453952
	gndr _||_ mdfn	0.038072388	3.345498e-05	0.020092379	0.05602782
Marital _status	Ad_S _||_ Cr_Q | mdfn	-0.039868616	1.397060e-05	-0.05782171	-0.021889789
	Ad_S _||_ crsp | mdfn	-0.006234368	4.971488e-01	-0.02422618	0.011761483
	Ad_S _||_ mrtl	-0.007787891	3.963165e-01	-0.02577805	0.010207313
	crsp _||_ mrtl	-0.010251827	2.641678e-01	-0.02824039	0.007743375
	mrtl _||_ mdfn	0.054580624	2.674182e-09	0.03662282	0.072503325
State_id	Ad_S _||_ Cr_Q | mdfn	-0.039868616	1.397060e-05	-0.05782171	-0.021889789
	Ad_S _||_ crsp | mdfn	-0.006234368	4.971488e-01	-0.02422618	0.011761483
	Ad_S _||_ sttd	-0.037476254	4.435163e-05	-0.05543268	-0.019495627
	crsp _||_ sttd	-0.025144973	6.159131e-03	-0.04311925	-0.007154437
	mdfn _||_ sttd	0.022949507	1.242164e-02	0.00495778	0.040926388
Lga_id	Ad_S _||_ Cr_Q | mdfn	-0.039868616	1.397060e-05	-0.057821709	-0.021889789
	Ad_S _||_ crsp | mdfn	-0.006234368	4.971488e-01	-0.024226183	0.011761483
	Ad_S _||_ lgad	-0.037808666	3.791819e-05	-0.055764538	-0.019828382
	crsp _||_ lgad	-0.027174936	3.072125e-03	-0.045146646	-0.009185654
	lgad _||_ mdfn	0.021737056	1.789339e-02	0.003744747	0.039715303
Age	Ad_S _||_ Age	-0.012737162	1.653457e-01	-0.03072390	0.005257817
	Ad_S _||_ Cr_Q | mdfn	-0.039868616	1.397060e-05	-0.05782171	-0.021889789
	Ad_S _||_ crsp | mdfn	-0.006234368	4.971488e-01	-0.02422618	0.011761483
	Age _||_ crsp	-0.132160076	1.653910e-47	-0.14980269	-0.114439757
	Age _||_ mdfn	0.444718095	0.000000e+00	0.43122216	0.460516727

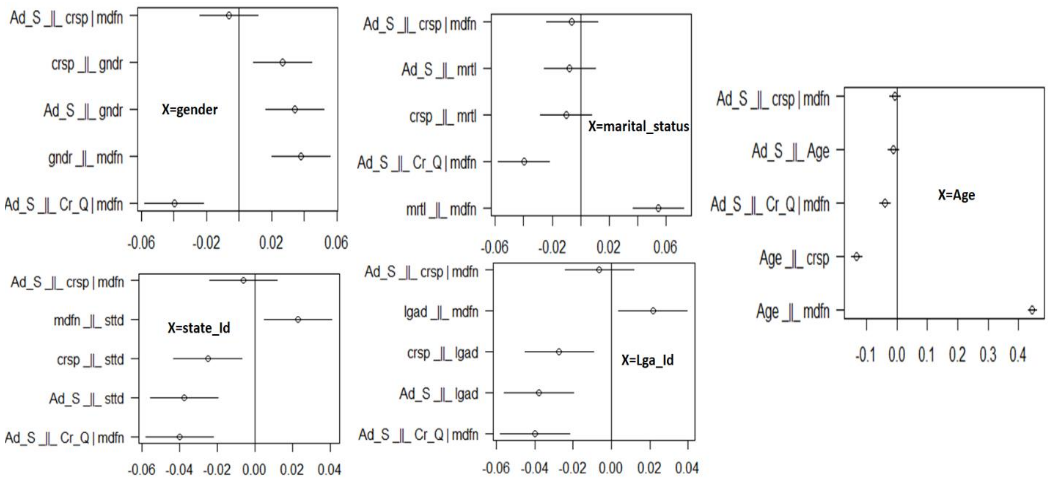

Figure 7. Results of the plotLocalTestResults function for the 5 X variable instances using equation 6 at 95% CI.

Algorithm 1. Computation of CIT Validation for Admission process for Polytechnic Education in Nigeria

INPUT: $X, CA, ME, CQ, AS, SCM\ CIT$
OUTPUT: $p - corrCoef, p.value, CI$
1: Start
2: Declare {X:= Set of individual characteristics features. Where $X \in \{x_1, x_2, \ldots x_n\}$
CA:= Course Applied.
ME:= Mode of Entry. Where $ME \in \{0,1\}$

Algorithm 1. Computation of CIT Validation for Admission process for Polytechnic Education in Nigeria

CQ:= Current Qualification. Where $CQ \in \{0,1\}$
AS:= Admission Status. Where AS is the label and a binary class
CI:= Confidence Interval @ 95%
$SCM\ CIT$:= CIT criteria identified in the SCM (encoded Assumptions)
P:= Probability of}

3: Read $X, CA, ME, CQ, AS, SCM\ CIT$.
4: for X: = x_1,
compute $\{P\ (SCM\ CIT)\}$
print $p - corrCoef,\ p.value,\ CI$
plot (print)
5: if $p - corrCoef\ =\ 0,\ p.value\ >\ 0.05, AND\ CI\ <=\ 0$ then
print "CIT validation confirmed"
else
print "CIT validation not confirmed"
6: for := $x_2, \dots x_n$
Repeat steps 3-5:
7: End

4.3. SCM Ontological Framework Validation Results Interpretation

The results of the SCM validation are evinced in Table 1 and Figure 7, and section 4.3.1 explicates its meaning.

4.3.1. Understanding the Results

The LocalTest function, a Digitty library in R programming, is employed in the CIT validation of the results. Its takes the coordinates and the CIT criteria identified in SCM ontological framework of equation 6 or 6b as inputs alongside the dataset distribution to perform the CIT. The outputs of the LocalTest are the Pearson correlation coefficient estimates, the p-value, and the confidence interval (CI) of the correlation coefficient for every conditional independence that is assumed in the SCM DAG. The variation of the Pearson correlation coefficient is between -1 and 1. Thus with 0 output implies the presence of no correlation, and a -1 or 1 indicates a perfect linear correlation. The outputted p-value for the LocalTest function shows the probability of observing the distribution of the dataset under the assumption that the independence condition holds. Thus, if the experiment outputs a correlation coefficient of around zero, with a high p-value (>0.05), it is interpreted that the assumptions of the implied conditional independence in the model structure hold true. On the other hand, a high value for the correlation coefficient, with a small or low p-value, evinced that the conditional independencies assumed in the SCM DAG ontological framework do not hold in the dataset distribution. Further, the 2.5% and 97% columns show the correlational coefficient's 95% confidence interval (CI). Thus, the smaller or narrower the CI is, and the further away from 0, evinces a negation of the conditional independencies assumed in the SCM ontological framework. Notably, the CIT of a SCM DAG requires that the associate effects of the identified conditional (in)dependencies be zero. Therefore, each effect that is numerically zero or close to it affirms the CIT assumptions encoded in the SCM DAG ontological framework, and it is considered the outcome of the desire for the validation process. However, the reverse is the case when the outcome is statistically significant – shifting significantly away from the zero mark. When this happens, the validation process is not considered successful because there is something wrong with the SCM DAG ontological framework designed or the dataset distribution has issues [62]–[64], [93].

Another function of the Dagitty library tool used in R for performing the test is the plotLocalTestResults function. As the name implies, this function plots graphically the result of the LocalTest function, enabling the accurate visualization and interpretation of the output results. Therefore, when the plotted graph coordinates line points to the zero line, it signifies or validates the assumptions in the DAG – meaning there is no dependency or correlation between the variables in focus. And the further away from the zero lines, the stronger is the evidence against the implied CIT assumptions in the SCM DAG model [93]. This function is also applied at a CI interval of 95% to show the uncertainty associated with the estimates.

4.3.2. SCM Ontology Results Interpretation

Thus, Table 1 and Figure 7 show the output results from the CIT, using the conditional independent assumptions of equation 6 or 6b, obtained from the SCM Direct Acyclic Graph (DAG) ontological framework of Figure 6. Since the variable set X in Figure 6 is only a representation of the 5 variables, which are gender, marital_status, state_Id, lga_id and Age variables, this CIT test uses the assumptions of equation 6, is performed 5 times for each instance of X, and the results is presented in Table 1 and Figure 7. From the results of the Pearson correlation coefficient estimates, it can be seen that all estimates of the coefficient are close to the zero mark, with a very narrow CI measured at 95%, and all p-values are above 0.05 (>0.05). Also, the plotLocalTestResults function shows the that the data distribution points are all align on the zero mark or closer to the zero. These results clearly validate the assumption evinced in equation 6 or 6b, which is concomitant to the validation if the SCM ontological framework for the admission process of Benpoly dataset.

4.4. LIME Explanations Presentation

This section presents the experiment result for LIME, where the LIME results produced fairness bias (section 4.41) and where the LIME fairness bias is constrained or ablated (section 4.4.2).

4.4.1. LIME Presentation Results with Fairness Bias

Figure 5 shows the dataset visualization after preprocessing and arriving at the nine (9) most important features in the admission process in the Polytechnic Education system in Nigeria. Eight of the nine features are considered input features, and one (Admission_status) is the outcome or targeted label. The prediction task is a binary classifier, where 1 depicts a prospective student who got admitted, and 0 depicts a prospective student who failed to be admitted. The black-box ML predictive algorithm employed is the Gaussian Naïve Based (GNB) classifier. The ML prediction accuracy is 89%, with an F1 score of 94%, and implementing the LIME explanation framework of Equations (2) and (3) with all the eight input features, the explanations results for the LIME individual feature weights as plotted in the graphs of Figure 8 for the eight inputs features are shown in Table 2. From the LIME explanations, the individual attributes features depicted as X in the SCM ontological framework of Figure 6 are known to have the highest weights (positively and negatively) that seem to influence the LIME explanations (i.e., X = {gender, martialstatus, stateId, lgaid, Age}). These LIME explanations for the Benpoly admission process (which is indeed an extrapolation for all admission processes in the Nigeria Polytechnic Education system) are obviously biased, as the results show that prospective students got admitted based on either their state of origin, sex, age, etc. These features are not the criteria for which a prospective student gains admission. Albeit, the LIME explanation framework made its explanation choices based on the patterns learned from the dataset (faithfulness or fidelity). Thus, these LIME explanations, as evinced, are not a true representation of the ground truth regarding the admission process in these institutions and, therefore, are not valid explanations for the admission process in the polytechnic admission process in Nigeria. These features, which are identified as having the highest weights, have nothing to do with the admission requirements in the polytechnic education system in Nigeria. Hence, there is a need for our fair-LIME explanation framework, which is implemented in the results of Table 3 and Figures 9 and 10 in section 4.4.2.

Table 2. Results of the LIME explanations on ML predictions for 5 data instances with fairness bias.

DI	LIME Feature Weight Explanations								Outcomes	
	SO	LGA	MS	ME	GEN	CA	CQ	AGE	Predicted	Actual
1	+0.38	+0.03	-0.03	+0.03	+0.02	+0.01	+0.00	+0.00	0.96	1
2	+0.41	+0.04	-0.03	+0.03	-0.02	+0.01	+0.01	+0.00	0.96	1
3	+0.39	+0.04	-0.02	-0.03	-0.01	-0.02	-0.01	+0.01	0.92	1
4	+0.40	+0.03	-0.03	+0.02	-0.02	0.00	-0.01	+0.01	0.96	1
5	-0.37	-0.12	-0.00	+0.02	+0.02	+0.00	+0.01	-0.03	0.00	0

DI – Data Instance.

The next narrative explicates the result outlined in Figure 8:

Subplot 1: LIME Explanation for Instance 1 - Title: "LIME Explanation for Admission

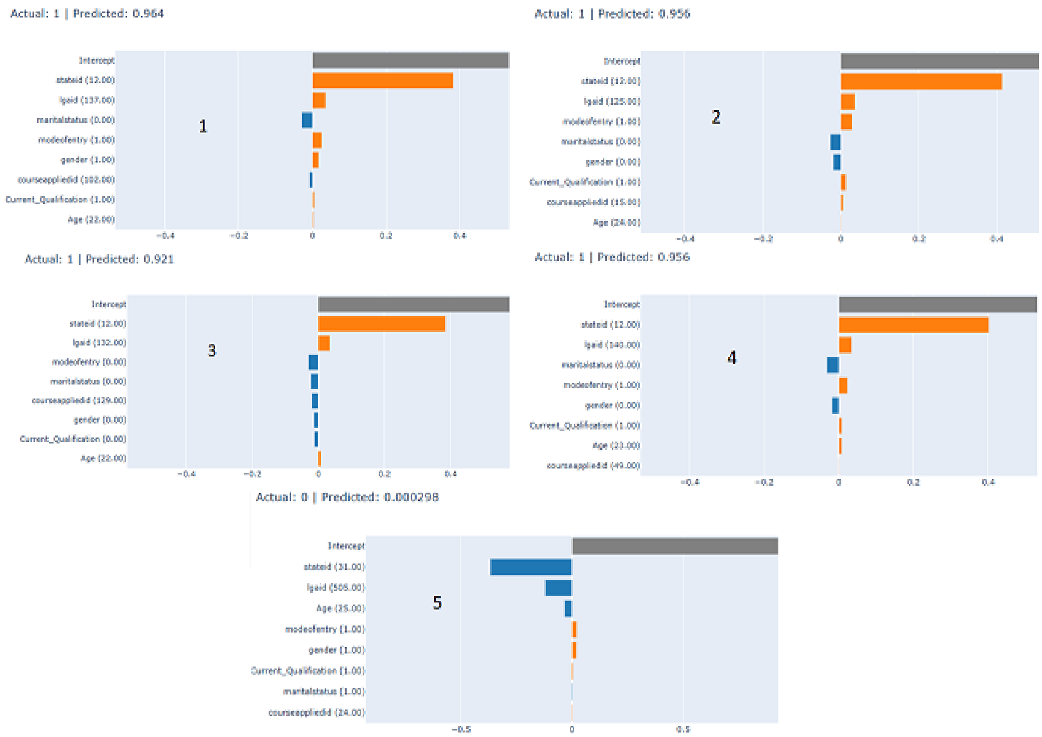

Figure 8. Visualized LIME Graph plot for Local_Explanations for five data instances on All identified features in the SCM Ontology.

Decision: Instance 1" - Description: "This subplot shows the LIME feature importance scores for the first instance in the test set. The applicant is predicted to be admitted (AS = 1) with a probability of 0.96. The top contributing features are State of Origin (SO), Local Government Area (LGA), and Mode of Entry (ME), while the other features have relatively smaller contributions."

Subplot 2: LIME Explanation for Instance 2 - Title: "LIME Explanation for Admission Decision: Instance 2" - Description: "This subplot presents the LIME feature importance scores for the second instance in the test set. The applicant is predicted to be admitted (AS = 1) with a probability of 0.96. Similar to Instance 1, the top contributing features are State of Origin (SO), Local Government Area (LGA), and Mode of Entry (ME), indicating a consistent pattern in the model's decision-making process."

Subplot 3: LIME Explanation for Instance 3 - Title: "LIME Explanation for Admission Decision: Instance 3" - Description: "This subplot displays the LIME feature importance scores for the third instance in the test set. The applicant is predicted to be admitted (AS = 1) with a probability of 0.92. The feature contributions follow a similar pattern to the previous instances, with State of Origin (SO), Local Government Area (LGA), and Mode of Entry (ME) having the highest impact on the prediction.".

Subplot 4: LIME Explanation for Instance 4 - Title: "LIME Explanation for Admission Decision: Instance 4" - Description: "This subplot illustrates the LIME feature importance scores for the fourth instance in the test set. The applicant is predicted to be admitted (AS = 1) with a probability of 0.96. The feature contributions are consistent with the other admitted instances, highlighting the influence of demographic factors like State of Origin (SO) and Local Government Area (LGA) on the model's decision.".

Subplot 5: LIME Explanation for Instance 5 - Title: "LIME Explanation for Admission Decision: Instance 5" - Description: "This subplot shows the LIME feature importance scores for the fifth instance in the test set. Unlike the previous instances, this applicant is predicted to be not admitted (AS = 0) with a probability of 1.00. The top contributing features are State of Origin (SO) and Local Government Area (LGA), but with negative importance scores, indicating that these factors are driving the non-admission decision.". Thus, from the explanations obtained from the automated LIME frameworks without our framework (i.e., constraining or ablating biased features), the LIME explanations are discovered to be biased in explaining important admission criteria. It only learned the patterns in the dataset, and made explanations on features such as state, gender, age, etc., which do not influence the admission selection criteria.

4.4.2. LIME Explanations Results with Ablated Input Features

To implement our fair-LIME explanation framework, the mutation of the SCM ontological framework of Figure 6 is imperative to perform the ablation process on the dataset. Thus, the five individual features identified and depicted as X, will need to be ablated from the SCM and concomitantly the dataset. This qualitative process is needed to identify, constrain, and sparse the features that can be inputted into the ML/LIME automated process to bring about predictions and LIME explanations with the dataset that are faithful and devoid of bias fairness bias. Hence, the fair-LIME SCM ontological framework is shown in Figure 9, as the X features are ablated or mutilated from the original SCM ontological framework of Figure 6. Thus, after the ablation process, our fair-LIME framework depicted in Figure 9, alongside constrained and sparse LIME Equations (4) and (5), will apply. Hence, the results for our fair-LIME framework results are shown in Figure 10 and Table 3. Therefore, instead of feeding the 8 input features into the automated LIME process, we now feed the LIME process with 3 input features (i.e., courseappliedid, modeofentry, and Current_Qualification) which are considered sine-quo-non for the admission process in Nigerian Polytechnic education system, as shown in Figure 9. The results from our fair-LIME framework results, as shown in Table 3 and Figure 10, has succeeded in removing the fairness bias in the initial LIME explanations results of Tables 2 and Figure 8. Thus, striking a balance between LIME explanations fidelity and fairness. Also, the black-box Gaussian Naïve based ML predictive model accuracy and F1 scores increased a bit to 90% and 95%, respectively.

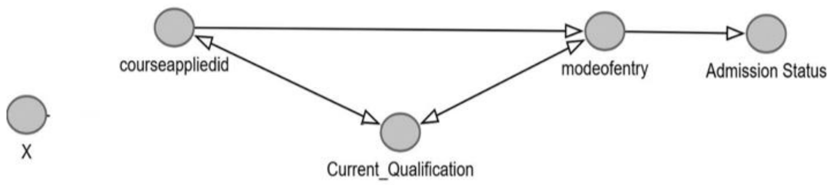

Figure 9. Ablated SCM Ontological Framework for BenPoly Admission, where the bias features set X is mutilated from the rest of the SCM features

Figure 10. Visualized LIME Graph plot for Local_Explanations for the 5 Data instance, with the Ablation Implementation on the SCM Ontology

Table 3. Result of the Ablation application on LIME explanations for 5 data instances

Data Instance	LIME Feature Weight Explanations			Outcomes	
	ME	CA	CQ	Predicted	Actual
1	+0.03	-0.01	+0.01	0.93	1
2	+0.03	+0.01	+0.01	0.94	1
3	-0.03	-0.01	-0.01	0.89	1
4	+0.03	+0.01	-0.00	0.94	1
5	+0.03	+0.01	+0.01	0.94	0

5. Conclusions

This study successfully applied the KDD process to extract and discover knowledge from Benpoly's admission database in Nigeria. An application-based SCM ontological framework was designed to represent the admission process in the Nigerian polytechnic education system, leveraging the SCM's ability to identify causal relations among features and validate the ontology's correctness using the CIT criteria.

The SCM ontology was employed to identify features causing fairness bias in the automated ML predictions and LIME framework. By constraining and ablating these biased features, the proposed fair-LIME framework produced more stable and fair explanations compared to the original LIME framework, with improved prediction accuracy (91% vs. 89%) and F1 scores (95% vs. 94%).

The main contribution of this study is the novel integration of a qualitative SCM ontology with quantitative ML and LIME methods to enforce fairness in explanations. This approach addresses the problem of biased explanations generated by automated ML and LIME frameworks when based on patterns learned from datasets that violate fairness. The fair-LIME framework demonstrates the importance of incorporating domain knowledge through ontologies to identify and mitigate biases in ML explanations.

However, a limitation of the fair-LIME framework is its context-specificity, as the SCM ontology is derived from the background knowledge of a particular process. Therefore, the framework can only be extrapolated to similar contexts, and different contexts would require modeling new SCM ontologies for each process before applying the fair-LIME framework.

For future work, comparing the performance of other explanation frameworks, such as Shapley explanations, with the fair-LIME framework on the same dataset could provide valuable insights into their relative strengths and weaknesses. Additionally, exploring the application of the fair-LIME framework to other domains and datasets with different fairness challenges would further demonstrate its generalizability and potential impact on promoting fairness in ML explanations.

Author Contributions: Author 1 conceptualized the study, conducted the experiments, analyzed the results, and prepared the initial draft of the manuscript. Authors 2, 3 and 4 provided guidance and supervision throughout the research process, validated the experimental methodology, reviewed, and edited the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The source code used in this study can be accessed at GitHub Repository: https://github.com/igobenny/Codes-SCM-LIME-ML-Explanations.git

Acknowledgments: We acknowledge Benue State Polytechnic, Ugbokolo Benue State, Nigeria for granting access to their admission database, which was essential for this study.

Conflicts of Interest: The authors declare no conflict of interest and have no financial or proprietary interests in any of the materials, methods, or findings discussed in this work.

References

[1] T. Calders and B. Custers, "What Is Data Mining and How Does It Work?," in *Discrimination and Privacy in the Information Society*, vol. 3, B. Custers, T. Calders, B. Schermer, and T. Zarsky, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 27–42. doi: 10.1007/978-3-642-30487-3_2.

[2] L. Cao and C. Zhang, "The evolution of KDD: Towards domain-driven data mining," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 21, no. 04, pp. 677–692, Jun. 2007, doi: 10.1142/S0218001407005612.

[3] M.-S. Chen, J. Han, and P. S. Yu, "Data mining: an overview from a database perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 8, no. 6, pp. 866–883, 1996, doi: 10.1109/69.553155.

[4] J. W. Seifert, "Data Mining: An Overview," 2004.

[5] D. T. Larose and C. D. Larose, *Discovering knowledge in data*, 2nd ed. Hoboken, NJ: Wiley-Blackwell, 2014.

[6] N. Ye, *The Handbook of Data Mining*. CRC Press, 2003. doi: 10.1201/b12469.

[7] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Upper Saddle River, NJ: Pearson, 2005.

[8] C. Rygielski, J.-C. Wang, and D. C. Yen, "Data mining techniques for customer relationship management," *Technol. Soc.*, vol. 24, no. 4, pp. 483–502, Nov. 2002, doi: 10.1016/S0160-791X(02)00038-6.

[9] G. S. Linoff and M. J. A. Berry, *Data mining techniques*, 3rd ed. Nashville, TN: John Wiley & Sons, 2011.

[10] O. P. Rud, *Data mining cookbook*. Nashville, TN: John Wiley & Sons, 2000.

[11] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2592–2602, Mar. 2009, doi: 10.1016/j.eswa.2008.02.021.

[12] F. Guo and H. Qin, "Data Mining Techniques for Customer Relationship Management," *J. Phys. Conf. Ser.*, vol. 910, p. 012021, Oct. 2017, doi: 10.1088/1742-6596/910/1/012021.

[13] P. S. Raju, V. R. Bai, and G. K. Chaitanya, "Data mining: Techniques for Enhancing Customer Relationship Management in Banking and Retail Industries," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 2, no. 1, pp. 2650–2657, 2014.

[14] R. Dass, "Data Mining in Banking and Finance: A Note for Bankers." Indian Institute of Management Ahmadabad, 2006.

[15] V. Jayasree and R. V. S. Balan, "A Review on Data Mining in Banking Sector," *Am. J. Appl. Sci.*, vol. 10, no. 10, pp. 1160–1165, Oct. 2013, doi: 10.3844/ajassp.2013.1160.1165.

[16] K. I. Moin and Q. B. Ahmed, "Use of data mining in banking," *Int. J. Eng. Res. Appl.*, vol. 2, no. 2, pp. 738–742, 2012.

[17] N. K. Hariharan, "Applications of data mining in finance," *Int. J. Innov. Eng. Res. Technol.*, vol. 5, no. 2, pp. 72–77, 2018.

[18] I. Yoo *et al.*, "Data Mining in Healthcare and Biomedicine: A Survey of the Literature," *J. Med. Syst.*, vol. 36, no. 4, pp. 2431–2448, Aug. 2012, doi: 10.1007/s10916-011-9710-5.

[19] N. Jothi, N. A. Rashid, and W. Husain, "Data Mining in Healthcare – A Review," *Procedia Comput. Sci.*, vol. 72, pp. 306–313, 2015, doi: 10.1016/j.procs.2015.12.145.

[20] H. Kaur and S. K. Wasan, "Empirical Study on Applications of Data Mining Techniques in Healthcare," *J. Comput. Sci.*, vol. 2, no. 2, pp. 194–200, Feb. 2006, doi: 10.3844/jcssp.2006.194.200.

[21] S. M. Birjandi and S. H. Khasteh, "A survey on data mining techniques used in medicine," *J. Diabetes Metab. Disord.*, vol. 20, no. 2, pp. 2055–2071, Aug. 2021, doi: 10.1007/s40200-021-00884-2.

[22] R. Bellazzi, F. Ferrazzi, and L. Sacchi, "Predictive data mining in clinical medicine: a focus on selected methods and applications," *WIREs Data Min. Knowl. Discov.*, vol. 1, no. 5, pp. 416–430, Sep. 2011, doi: 10.1002/widm.23.

[23] J. A. Harding, M. Shahbaz, Srinivas, and A. Kusiak, "Data Mining in Manufacturing: A Review," *J. Manuf. Sci. Eng.*, vol. 128, no. 4, pp. 969–976, Nov. 2006, doi: 10.1115/1.2194554.

[24] Y. Guo, W. Zhang, Q. Qin, K. Chen, and Y. Wei, "Intelligent manufacturing management system based on data mining in artificial intelligence energy-saving resources," *Soft Comput.*, vol. 27, no. 7, pp. 4061–4076, Apr. 2023, doi: 10.1007/s00500-021-06593-5.

[25] S. Kamble, A. Desai, and P. Vartak, "Data mining and data warehousing for Supply Chain Management," in *2015 International Conference on Communication, Information & Computing Technology (ICCICT)*, Jan. 2015, pp. 1–6. doi: 10.1109/ICCICT.2015.7045692.

[26] A. Kusiak, "Data mining: manufacturing and service applications," *Int. J. Prod. Res.*, vol. 44, no. 18–19, pp. 4175–4191, Sep. 2006, doi: 10.1080/00207540600632216.

[27] M. Er Kara, S. Ü. Oktay Fırat, and A. Ghadge, "A data mining-based framework for supply chain risk management," *Comput. Ind. Eng.*, vol. 139, p. 105570, Jan. 2020, doi: 10.1016/j.cie.2018.12.017.

[28] M. Wu, K. Liu, and H. Yang, "Supply chain production and delivery scheduling based on data mining," *Cluster Comput.*, vol. 22, no. S4, pp. 8541–8552, Jul. 2019, doi: 10.1007/s10586-018-1894-8.

[29] V. R. Rao, "A Framework for e-Government Data Mining Applications (eGDMA) for Effective Citizen Services-An Indian Perspective," *Int. J. Comput. Sci. Inf. Technol. Res.*, vol. 2, no. 4, pp. 209–225, 2014.

[30] M. M. Mostafa and A. A. El-Masry, "Citizens as consumers: Profiling e-government services' users in Egypt via data mining techniques," *Int. J. Inf. Manage.*, vol. 33, no. 4, pp. 627–641, Aug. 2013, doi: 10.1016/j.ijinfomgt.2013.03.007.

[31] S. De Cnudde and D. Martens, "Loyal to your city? A data mining analysis of a public service loyalty program," *Decis. Support Syst.*, vol. 73, pp. 74–84, May 2015, doi: 10.1016/j.dss.2015.03.004.

[32] S. Ramos and Z. Vale, "Data mining techniques application in power distribution utilities," in *2008 IEEE/PES Transmission and Distribution Conference and Exposition*, Apr. 2008, pp. 1–8. doi: 10.1109/TDC.2008.4517229.

[33] D. M. de Santana, S. R. Lourenço, and D. A. Cassiano, "Data mining approach for energy efficiency improvements in a utilities supply on a petrochemical plant," *Evol. Syst.*, vol. 14, no. 6, pp. 1071–1081, Dec. 2023, doi: 10.1007/s12530-023-09515-y.

[34] R. R, E. S. Sankari, and Matheswaran.P, "Detection of non-technical loss in power utilities using data mining techniques," *Int. J. Innov. Res. Sci. Technol.*, vol. 1, no. 9, pp. 97–101, 2015.

[35] S. Singh and A. Yassine, "Big Data Mining of Energy Time Series for Behavioral Analytics and Energy Consumption Forecasting," *Energies*, vol. 11, no. 2, p. 452, Feb. 2018, doi: 10.3390/en11020452.

[36] S. Ramos *et al.*, "Data mining techniques for electricity customer characterization," *Procedia Comput. Sci.*, vol. 186, pp. 475–488, 2021, doi: 10.1016/j.procs.2021.04.168.

[37] L. Zhu, M. Li, Z. Zhang, X. Du, and M. Guizani, "Big Data Mining of Users' Energy Consumption Patterns in the Wireless Smart Grid," *IEEE Wirel. Commun.*, vol. 25, no. 1, pp. 84–89, Feb. 2018, doi: 10.1109/MWC.2018.1700157.

[38] C. Costa and M. Y. Santos, "Improving cities sustainability through the use of data mining in a context of big city data," in *International Conference of Data Mining and Knowledge Engineering*, 2015. [Online]. Available: https://www.iaeng.org/publication/WCE2015/WCE2015_pp320-325.pdf

[39] C. Cacciuttolo, V. Guzmán, P. Catriñir, E. Atencio, S. Komarizadehasl, and J. A. Lozano-Galant, "Low-Cost Sensors Technologies for Monitoring Sustainability and Safety Issues in Mining Activities: Advances, Gaps, and Future Directions in the Digitalization for Smart Mining," *Sensors*, vol. 23, no. 15, p. 6846, Aug. 2023, doi: 10.3390/s23156846.

[40] Y. Shi and U. Zhu, "Efficient Data Mining Method for Environmental Monitoring around Water Conservancy Projects," *Ekoloji Derg.*, vol. 28, no. 108, pp. 2447–2451, 2019.

[41] K. Gibert, J. Izquierdo, M. Sànchez-Marrè, S. H. Hamilton, I. Rodríguez-Roda, and G. Holmes, "Which method to use? An assessment of data mining methods in Environmental Data Science," *Environ. Model. Softw.*, vol. 110, pp. 3–27, Dec. 2018, doi: 10.1016/j.envsoft.2018.09.021.

[42] R. S. J. d. Baker, "Data Mining," in *International Encyclopedia of Education*, Elsevier, 2010, pp. 112–118. doi: 10.1016/B978-0-08-044894-7.01318-X.

[43] C. Romero and S. Ventura, "Data mining in education," *WIREs Data Min. Knowl. Discov.*, vol. 3, no. 1, pp. 12–27, Jan. 2013, doi: 10.1002/widm.1075.

[44] A. Algarni, "Data Mining in Education," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 6, 2016, doi: 10.14569/IJACSA.2016.070659.

[45] K. R. Koedinger, S. D'Mello, E. A. McLaughlin, Z. A. Pardos, and C. P. Rosé, "Data mining and education," *WIREs Cogn. Sci.*, vol. 6, no. 4, pp. 333–353, Jul. 2015, doi: 10.1002/wcs.1350.

[46] S. Agarwal, "Data Mining in Education: Data Classification and Decision Tree Approach," *Int. J. e-Education, e-Business, e-Management e-Learning*, vol. 2, no. 2, pp. 140–144, 2012, doi: 10.7763/IJEEEE.2012.V2.97.

[47] M. M. Arcinas, una S. Sajja, S. Asif, S. Gour, E. Okoronkwo, and M. Naved, "Role of data mining in education for improving students performance for social change," *Turkish J. Physiother. Rehabil.*, vol. 32, no. 3, pp. 6519–6526, 2021.

[48] S. Suhirman, T. Herawan, H. Chiroma, and J. Mohamad Zain, "Data Mining for Education Decision Support: A Review," *Int. J. Emerg. Technol. Learn.*, vol. 9, no. 6, p. 4, Dec. 2014, doi: 10.3991/ijet.v9i6.3950.

[49] M. Goyal and R. Vohra, "Applications of Data Mining in Higher Education," *Int. J. Comput. Sci. Issues*, vol. 9, no. 2, pp. 113–120, 2012.

[50] S. K. Malik, N. Prakash, and S. Rizvi, "Developing an university ontology in education domain using protégé for semantic web," *Int. J. Eng. Sci. Technol.*, vol. 2, no. 9, pp. 4673–4681, 2010.

[51] A. Ameen, K. R. Khan, and B. P. Rani, "Construction of university ontology," in *2012 World Congress on Information and Communication Technologies*, Oct. 2012, pp. 39–44. doi: 10.1109/WICT.2012.6409047.

[52] N. Malviya, N. Mishra, and S. Sahu, "Developing University Ontology using protégé OWL Tool: Process and Reasoning," *Int. J. Sci. Eng. Res.*, vol. 2, no. 9, 2011.

[53] L. Zeng, T. Zhu, and X. Ding, "Study on Construction of University Course Ontology: Content, Method and Process," in *2009 International Conference on Computational Intelligence and Software Engineering*, Dec. 2009, pp. 1–4. doi: 10.1109/CISE.2009.5363158.

[54] I. T. Ayorinde and O. A. Akinyele, "A Formal Activity Ontology for Postgraduate Admission Processes (A Case Study of University of Ibadan, Nigeria)," *Ilorin J. Comput. Sci. Inf. Technol.*, vol. 1, no. 1, pp. 1–18, 2016.

[55] G. T. Ayem, S. G. Thandekkattu, A. S. Nsang, and M. Fonkam, "Structural Causal Model Design and Causal Impact Analysis: A Case of SENSE-EGRA Dataset," 2023, pp. 39–52. doi: 10.1007/978-981-99-3485-0_4.

[56] G. T. Ayem, A. Ajibesin, A. Iorliam, and A. S. Nsang, "A mixed framework for causal impact analysis under confounding and selection biases: a focus on Egra dataset," *Int. J. Inf. Technol.*, Oct. 2023, doi: 10.1007/s41870-023-01490-6.

[57] G. T. Ayem, A. S. Nsang, B. I. Igoche, and G. Naankang, "Design and Validation of Structural Causal Model: A focus on SENSE-EGRA Datasets," *Int. J. Adv. Sci. Comput. Eng.*, vol. 5, no. 3, pp. 257–268, Dec. 2023, doi: 10.62527/ijasce.5.3.177.

[58] G. T. Ayem, O. Asilkan, and A. Iorliam, "Design and Validation of Structural Causal Model: A Focus on EGRA Dataset," *J. Comput. Theor. Appl.*, vol. 1, no. 2, pp. 86–103, Nov. 2023, doi: 10.33633/jcta.v1i2.9304.

[59] S. Sheikholeslami, "Ablation Programming for Machine Learning," KTH-Royal Institute of Technology, 2019.

[60] S. Sheikholeslami, M. Meister, T. Wang, A. H. Payberah, V. Vlassov, and J. Dowling, "AutoAblation," in *Proceedings of the 1st Workshop on Machine Learning and Systems*, Apr. 2021, pp. 55–61. doi: 10.1145/3437984.3458834.

[61] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?,'" in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[62] J. Pearl, *Causality: Models, Reasoning and Inference*. University of California, Los Angeles: Cambridge University Press, 2009.

[63] S. Greenland, J. Pearl, and J. M. Robins, "Causal diagrams for epidemiologic research.," *Epidemiology*, vol. 10, no. 1, pp. 37–48, Jan. 1999, [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9888278

[64] J. Pearl, "Causal inference in statistics: An overview," *Stat. Surv.*, vol. 3, no. none, Jan. 2009, doi: 10.1214/09-SS057.

[65] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Stat.*, vol. 32, no. 2, Apr. 2004, doi: 10.1214/009053604000000067.

[66] J. Budzianowski *et al.*, "Machine learning model for predicting late recurrence of atrial fibrillation after catheter ablation," *Sci. Rep.*, vol. 13, no. 1, p. 15213, Sep. 2023, doi: 10.1038/s41598-023-42542-y.

[67] C. K. Yeh, C. Y. Hsieh, A. S. Suggala, D. I. Inouye, and P. Ravikumar, "On the (In)fidelity and sensitivity of explanations," *Adv. Neural Inf. Process. Syst.*, vol. 32, no. NeurIPS, 2019, [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/a7471fdc77b3435276507cc8f2dc2569-Paper.pdf

[68] A. Papenmeier, G. Englebienne, and C. Seifert, "How model accuracy and explanation fidelity influence user trust in AI," in *IJCAI 2019 Workshop on Explainable Artificial Intelligence*, 2019.

[69] R. Gaudel, L. Galárraga, J. Delaunay, L. Rozé, and V. Bhargava, "s-LIME: Reconciling Locality and Fidelity in Linear Explanations," in *Advances in Intelligent Data Analysis XX*, 2022, pp. 102–114. doi: 10.1007/978-3-031-01333-1_9.

[70] M. Velmurugan, C. Ouyang, C. Moreira, and R. Sindhgatta, "Developing a Fidelity Evaluation Approach for Interpretable Machine Learning." Jun. 15, 2021. [Online]. Available: http://arxiv.org/abs/2106.08492

[71] A. Angerschmid, J. Zhou, K. Theuermann, F. Chen, and A. Holzinger, "Fairness and Explanation in AI-Informed Decision Making," *Mach. Learn. Knowl. Extr.*, vol. 4, no. 2, pp. 556–579, Jun. 2022, doi: 10.3390/make4020026.

[72] J. Dodge, Q. V. Liao, Y. Zhang, R. K. E. Bellamy, and C. Dugan, "Explaining models," in *Proceedings of the 24th International Conference on Intelligent User Interfaces*, Mar. 2019, pp. 275–285. doi: 10.1145/3301275.3302310.

[73] A. Balagopalan, H. Zhang, K. Hamidieh, T. Hartvigsen, F. Rudzicz, and M. Ghassemi, "The Road to Explainability is Paved with Bias: Measuring the Fairness of Explanations," in *2022 ACM Conference on Fairness, Accountability, and Transparency*, Jun. 2022, pp. 1194–1206. doi: 10.1145/3531146.3533179.

[74] E. A. Murray, "What have ablation studies told us about the neural substrates of stimulus memory?," *Semin. Neurosci.*, vol. 8, no. 1, pp. 13–22, Feb. 1996, doi: 10.1006/smns.1996.0003.

[75] B. Zhou, Y. Sun, D. Bau, and A. Torralba, "Revisiting the Importance of Individual Units in CNNs via Ablation." Jun. 07, 2018. [Online]. Available: http://arxiv.org/abs/1806.02891

[76] R. Meyes, M. Lu, C. W. de Puiseau, and T. Meisen, "Ablation Studies in Artificial Neural Networks." Jan. 24, 2019. [Online]. Available: http://arxiv.org/abs/1901.08644

[77] J. Pearl, "The seven tools of causal inference, with reflections on machine learning," *Commun. ACM*, vol. 62, no. 3, pp. 54–60, Feb. 2019, doi: 10.1145/3241036.

[78] C. Cinelli, D. Kumor, B. Chen, J. Pearl, and E. Bareinboim, "Sensitivity Analysis of Linear Structural Causal Models," in *Proceedings of the 36th International Conference on Machine Learning*, 2019, pp. 1252–1261. [Online]. Available: https://proceedings.mlr.press/v97/cinelli19a.html

[79] K. S. Gill, "Pearl, Judea and Mackenzie, Dana: The book of why: the new science of cause and effect (2018)," *AI Soc.*, vol. 35, no. 3, pp. 767–768, Sep. 2020, doi: 10.1007/s00146-020-00971-7.

[80] K. A. Markus, "Causal effects and counterfactual conditionals: contrasting Rubin, Lewis and Pearl," *Econ. Philos.*, vol. 37, no. 3, pp. 441–461, Nov. 2021, doi: 10.1017/S0266267120000437.

[81] R. Briggs, "Interventionist counterfactuals," *Philos. Stud.*, vol. 160, no. 1, pp. 139–166, Aug. 2012, doi: 10.1007/s11098-012-9908-5.

[82] A.-H. Karimi, B. Schölkopf, and I. Valera, "Algorithmic Recourse," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Mar. 2021, pp. 353–362. doi: 10.1145/3442188.3445899.

[83] F. Wang and C. Rudin, "Falling Rule Lists," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, 2015, vol. 38, pp. 1013–1022. [Online]. Available: https://proceedings.mlr.press/v38/wang15a.html

[84] J. Dai, S. Upadhyay, U. Aivodji, S. H. Bach, and H. Lakkaraju, "Fairness via Explanation Quality," in *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, Jul. 2022, pp. 203–214. doi: 10.1145/3514094.3534159.

[85] Y. Zhao, Y. Wang, and T. Derr, "Fairness and Explainability: Bridging the Gap towards Fair Model Explanations," *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 9, pp. 11363–11371, Jun. 2023, doi: 10.1609/aaai.v37i9.26344.

[86] J. Schoeffer and N. Kuehl, "Appropriate Fairness Perceptions? On the Effectiveness of Explanations in Enabling People to Assess the Fairness of Automated Decision Systems," in *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*, Oct. 2021, pp. 153–157. doi: 10.1145/3462204.3481742.

[87] U. Bhatt *et al.*, "Explainable machine learning in deployment," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Jan. 2020, pp. 648–657. doi: 10.1145/3351095.3375624.

[88] S. Sharma, J. Henderson, and J. Ghosh, "CERTIFAI: A common framework to provide explanations and analyse the fairness and robustness of black-box models," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Feb. 2020, pp. 166–172. doi: 10.1145/3375627.3375812.

[89] S. Tang *et al.*, "Machine Learning–Enabled Multimodal Fusion of Intra-Atrial and Body Surface Signals in Prediction of Atrial Fibrillation Ablation Outcomes," *Circ. Arrhythmia Electrophysiol.*, vol. 15, no. 8, Aug. 2022, doi: 10.1161/CIRCEP.122.010850.

[90] J. C. Miller, *Laser Ablation Principles and Applications*, vol. 28. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994. doi: 10.1007/978-3-642-78720-1.

[91] G. T. Ayem, O. Asilkan, A. Iorliam, R. Ibrahim, and S. George, "Causal Inference Estimates with Backdoor Adjustment Condition vs. the Unconfoundedness Assumption: A Comparative Analysis Study of the Structural Causal Model and the Potential Outcome Frameworks," *ESP J. Eng. Technol. Adv.*, 2023, doi: 10.56472/25832646/JETA-V3I8P101.

[92] R. U. Onyekwelu and A. J. Obikeze, "E-Administration And Service Delivery: A Study Of Joint Admissions And Matriculation Board (JAMB) Nigeria," *Int. J. Innov. Dev. Policy Stud.*, 2023, [Online]. Available: http://seahipaj.org/journals-ci/mar-2023/IJIDPS/full/IJIDPS-M-5-2023.pdf

[93] A. Ankan, I. M. N. Wortel, and J. Textor, "Testing Graphical Causal Models Using the R Package 'dagitty,'" *Curr. Protoc.*, vol. 1, no. 2, Feb. 2021, doi: 10.1002/cpz1.45.