Frontiers in Health Informatics
ISSN-Online: 2676-7104
www.healthinformaticsjournal.com
2024; Vol 13: Issue 8
Open Access

Explainable AI Models for Enhanced Decision-Making in Cybersecurity

[1]Shambhu Sharan Srivastava, [2]Rajalakshmi C N, [3]Komal Baburao Umare, [4]S. B G Tilak Babu, [5]Uruj Jaleel, [6]G. Muthupansi

[1]Shambhu Sharan Srivastava, Associate Professor, Department Of Computer Science, School Of Management Sciences, Varanasi, Bacchaon, Varanasi, Shambhuss@Yahoo.Com
[2]Rajalakshmi C N, Assistant Professor, Computerscience And Engineering, Mallareddy University, Mallareddy University Hyderabad,Rajecnphd@Gmail.Com
[3]Komal Baburao Umare, Assistant Professor, Electronics And Communication Engineering, Visvesvaraya National Institute Of Technology (VNIT), Nagpur, Ambazari, Nagpur, Maharashtra, Komal29umare@Gmail.Com
[4]S. B G Tilak Babu, Department Of ECE, Aditya Univercity, Surampalem, Thilaksayila@Gmail.Com
[5]Uruj Jaleel, Professor, Department Of MCA, Meerut Institute Of Engineering And Technology, Meerut,Meerut,Uttar Pradesh, Dr_Urujjaleel@Yahoo.Com
[6]G.Muthupansi, Assistant Professor, Computer Science And Engineering Chennai Institute Of Technology, Chennai, Pg.Muthupandi@Gmail.Com

Cite this paper as: Shambhu Sharan Srivastava, Rajalakshmi C N, Komal Baburao Umare, S. B G Tilak Babu, Uruj Jaleel, G. Muthupansi, (2024) Explainable AI Models for Enhanced Decision-Making in Cybersecurity. *Frontiers in Health Informatics*, *13(8) 1571-1577*

ABSTRACT
Enhancing decision-making with explainable artificial intelligence (XAI) models is essential to effective cybersecurity threat identification and mitigation as the complexity of cybersecurity threats continues to increase. The purpose of this research is to investigate how explainable artificial intelligence technologies might enhance decision-making in the field of cybersecurity. The chi-square test, which is used for feature selection, decision trees, which are used for classification, and data cleaning and normalization, which are used for preprocessing, the framework that we offer encompasses all three of these important methods. The process of cleaning and normalizing data ensures that raw cybersecurity data is put into a format that is consistent by resolving issues such as scale variances and missing values. Chi-Square is a statistical test that highlights the most statistically significant characteristics of a model, thereby reducing the complexity of the model and improving its interpretability. In conclusion, decision trees are utilized as the categorization model due to the fact that they are transparent by their very nature and offer choice paths that are easy for individuals to comprehend. This technique not only accomplishes high-performance detection, but it also helps cybersecurity experts comprehend model forecasts, contributes to the development of trust, and enables practical decision-making in situations where threats are occurring in real time.

Keywords: Explainable AI, Cybersecurity Decision-Making, Machine Learning Models, Data Cleaning, Normalization, Decision Trees, Model Interpretability

1. Introduction

Cybersecurity threats are becoming more complex and numerous, making data protection and system integrity difficult for enterprises. Rule-based systems and signature-based detection fail to keep up with complex threat trends. Machine learning (ML) algorithms' capacity to search massive data sets for unusual patterns and risks has stimulated cybersecurity ML model development. Machine learning models boost accuracy but are opaque, making decision-making harder for cybersecurity professionals [1]. This secrecy can damage confidence and make high-stakes decisions harder. This is why Explainable AI (XAI) models that provide explicit decision-making insights are crucial.

Explainable AI helps humans and complicated machine learning systems communicate. XAI improves decision-making, trust, and cybersecurity experts' understanding and use of AI model insights by giving interpretable explanations for model predictions. Our paper proposes an approach that uses machine learning to improve cybersecurity decision-making while maintaining model openness. In particular, we emphasize three methods: Chi-square test for feature selection, decision trees for classification, and data cleaning and normalization for preprocessing Cybersecurity data must be preprocessed before machine learning algorithms can use it. Cybersecurity datasets might

be noisy, inconsistent, or incomplete due to dynamic user activity, network traffic, and system logs. Missing numbers, mistakes, and incomplete datasets are addressed by data cleansing. Normalization is used to guarantee that all features are on a comparable scale because features with wide ranges can skew the model's output[2]. Normalization enhances model efficiency by ensuring that all characteristics contribute equally to learning and preventing prominent features from overshadowing others. Raw data must be preprocessed to increase explainability and model performance for machine learning.

Feature selection, which determines which features are most relevant to the model, follows data preparation. The Chi-Square test measures a characteristic's independence from the target variable. This test removes unnecessary features, lowering data dimensionality and improving model interpretability [3]. In cybersecurity, where the dataset may have many features, the Chi-Square test helps identify the most significant parameters that affect model prediction. The final model is practical and understandable.

Finally, we categorize using Decision Trees, a famous machine learning method with interpretability and openness [4]. Decision trees divide data by feature values using simple, clear rules. Due to their interpretability, decision trees are ideal for cybersecurity applications where model projection logic is crucial to operational decision-making [5]. Following a prediction through the tree structure helps cybersecurity professionals grasp the model's reasoning, making validation and action easier.

Finally, data cleansing and normalization, chi-square feature selection, and decision trees may improve cybersecurity AI model explainability and effectiveness. Focusing on these strategies, we hope to create a model that properly detects dangers and provides clear explanations, helping cybersecurity experts to make better judgments.

2. RELATED WORKS

Recent cybersecurity attention has centered on AI and machine learning. Cyberattacks' number and sophistication overwhelm traditional security solutions. Machine learning can detect anomalies, predict threats, and IDS. Understanding these models for real-world use is tough. XAI cybersecurity model data pretreatment, feature selection, and classification are covered here.

Data prep is needed for cybersecurity machine learning. Noise, incomplete, and unstructured cybersecurity datasets need pretreatment for model accuracy and efficiency. Cybersecurity requires data cleaning, according to many research. L. Breiman, “Random forests. (2001) [6] recommend removing unneeded data from anomaly-based intrusion detection machine learning models. Their investigation used several data cleaning methods to ensure uniformity and minimize errors from missing or erroneous records.

Normalisation ensures all features contribute equally to the model. A. Roy, R. K. K. Sahoo, and A. Abraham (2020) [7] recommend normalizing network traffic data, which often contains byte counts and packet sizes, to prevent one feature from dominating learning. Min-Max scaling and Z-score normalize cybersecurity data. On cybersecurity datasets, M. Chen, S. Mao, and Y. Liu (2014) [8] observed that normalization increases SVM and Decision Tree convergence.

Feature selection affects cybersecurity datasets with high-dimensional feature spaces' machine learning model efficiency and interpretability. Lowering dimensionality increases model performance and interpretability, according R. Caruana et al. (2015) [9] use Chi-Square to detect intrusions. Their research shows that the Chi-Square test can identify a dataset's most essential features by examining their statistical dependency on the target variable. Real-time cybersecurity applications benefit from Chi-Square feature selection's categorization accuracy and computer simplicity.

Chi-Square testing reduces wasted data from imbalanced class datasets in cybersecurity. This was examined by R. E. Schapire. (2003) [10]. They used statistical tests like Chi-Square to identify the most important variables to enhance intrusion detection rates while staying simple and explainable. Highlighting essential traits makes the model interpretable, boosting cybersecurity decision-making system trust.

Cybersecurity uses decision trees to comprehend numerical and categorical data. Because of their simplicity and clarity, decision trees are appropriate for cybersecurity applications that require forecast logic competence. Chandola et al. (2009) argue decision trees help IDS classify network traffic as regular or abnormal. Security analysts can investigate numerous decision paths using the decision tree model to break data into feature values.

Cybersecurity decision trees can distinguish regular and malicious network anomalies. Cybersecurity instances with defined decision-making procedures using C4.5 and CART algorithms. We presented Random Forests, an ensemble of decision trees more accurate than a single tree and transparent through feature importance analysis.

Decision trees are useful but can overfit noisy or high-dimensional data. Advised pruning decision trees to prevent overfitting and retain interpretability. This trimming method increases cybersecurity model generalization in noisy, uncertain situations.
In cybersecurity, machine learning model explainability research is crucial. Models show cybersecurity specialists' threat detection and mitigation choices. Explain black-box model predictions, including decision trees, using Local Interpretable Model-agnostic Explanations. Security analysts learn from LIME's comprehensive cybersecurity model predictions why blocking an IP address or flagging an anomaly is advised.
Another cybersecurity explainability method is SHAP. Showed how SHAP values might explain certain cybersecurity threat identification ensemble model predictions like Random Forests or Gradient Boosting Machines. SHAP provides a consistent framework to assign feature significance and understand how each item affects model decision to promote trust and accountability.
Preprocessing, feature selection, and explainable classification models are essential to cybersecurity AI system development. Data cleansing, normalization, and feature selection are recent research topics, with the Chi-Square test promising for cybersecurity dataset feature evaluation. Open and interpretable Decision Trees are popular cybersecurity categorization tools. This and explainable AI technologies like LIME and SHAP may improve cybersecurity AI model accuracy and trustworthiness over time.

3. RESEARCH METHODOLOGY

An explainable AI cybersecurity decision-making framework improves threat detection and model transparency and interpretability. Clean and normalize data, pick features with Chi-Square, then classify with Decision Trees as shown in Figure 1. Research methods include data collection, preprocessing, feature selection, model creation, and evaluation. As the AI model improves accuracy and explainability, cybersecurity pros can use it.

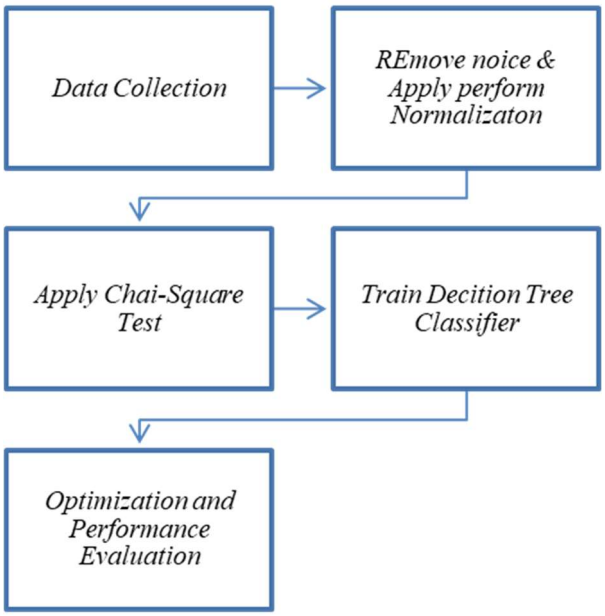

Figure 1: Shows Flow diagram for the proposed methodology.

3.1 Data Collection

We begin our research with cybersecurity data. Public cybersecurity datasets KDD Cup 1999 and CICIDS 2017 provide network traffic logs, intrusion data, and attack data. Popular cybersecurity datasets with normal and attack-related network traffic patterns are ideal for intrusion detection model construction.
Raw data includes continuous numerical values, category data, and time-series data [11]. Raw datasets must be turned into machine learning algorithm-friendly formats while keeping explainability.

3.2 Clean and normalize data

Preprocessing data for machine learning follows. Before machine learning, cybersecurity datasets must be cleaned of missing values, inconsistencies, and noise[12]. This research needs following preprocessing steps:

3.2.1 Data Cleaning

Network issues, sensor failures, and incomplete logs might cause raw dataset missing values. We use mean and mode imputation for numerical and categorical data. Missing data is handled carefully to avoid model bias. Remove rows with excessive missing data to protect data integrity. Outlier detection eliminates model-threatening data. Z-score and IQR identify outliers. The model's predictions are less impacted by rare, severe data following outlier removal or correction.

3.2.2 The normalization

Normalization is important to prevent one particular dataset parameter like network packet size, time length, or IP address counts from dominating model learning because to their different scales. The Min-Max normalization adjusts all numerical features to [0,1]. All features increase Decision Tree convergence and model performance equally.

3.3 Chi-Square Feature Selection

Features are needed for dimensionality reduction, model efficiency, and interpretability. High-dimensional datasets often have irrelevant or redundant characteristics, reducing model performance and explainability. This research selects characteristics using Chi-Square. The Chi-Square test analyzes features' independence from the target variable to determine categorization.

3.3.1 Chi-Square App

A dataset feature's statistical significance to the target variable is measured using the Chi-Square test. High Chi-Square results help model training, whereas low scores are discarded [13]. This stage reduces noise and improves model performance and interpretability using relevant data.

Features we use feature engineering and the Chi-Square test to build or change features based on domain expertise. Security indications like time of day and number of failed login attempts may be added. Engineering and selection are used in model training [14].

3.4 Class Decision Trees

Classification and interpretation are good with decision trees. Cybersecurity experts can grasp Decision Trees because they simplify complex decision-making into basic rules. Creating models CART uses classification and regression trees to generate decision trees. Recursively partitioning the dataset by the best homogenous subset separator creates binary trees in CART. Gini Impurity or Information Gain defines each tree node's ideal split [15]. Decision rules in the final tree structure classify network traffic as regular or malicious.

3.4.1 Tree trimming

Purging over fitted complicated decision trees decreases their size. Pruning branches that don't improve forecast accuracy. Clarity is achieved by pruning decision trees. After tree growth, we remove unneeded branches with least complexity.

Gini Index (Used in Decision Trees)

To measure impurity:

Gini=1− (p12+p22+⋯+pn2)

Where p is the probability of each class.

Information Gain

To calculate how much information is gained:

Information Gain=Entropy (Before)−Entropy (After)

Entropy

To measure uncertainty in the data:

Entropy=−p1log2(p1)−p2log2(p2)

Where p1 and p2 are class probabilities.

3.4.2 Model Explainability

Decision Trees are understood well. We use the tree to evaluate decision-making after model training. Each tree node represents a feature and threshold value, indicating how the model classifies input. Cybersecurity applications need transparency for model estimations and decisions.

3.5 Model Evaluation

Following training, we assess the model using machine learning metrics including accuracy, precision, recall, and F1-score. We report measures like ROC-AUC and Precision-Recall AUC to evaluate model performance in cybersecurity datasets with class imbalance, where assaults (positive class) are much less than typical instances (negative class).

We evaluate the model's explainability using decision tree visualization and feature importance. The model has high detection and simple prediction logic. Cybersecurity AI models are explainable after preprocessing, feature selection, and categorization. Normalizing and cleaning data prepares it for machine learning. Chi-Square tests prioritize attributes, while Decision Trees clarify the model. Give cybersecurity professionals accurate estimates and actionable insights into how they form to increase trust and decision-making in real-world cybersecurity scenarios.

4. RESULTS AND DISCUSSION

It has been demonstrated that the explainable artificial intelligence (XAI) technique that has been presented is capable of successfully enhancing decision-making in cybersecurity systems. This is accomplished through the utilization of data preprocessing, feature selection through the utilization of the Chi-square test, and classification through the utilization of Decision Trees. The phase of data preparation, which comprised data cleaning and normalization, resulted in a significant improvement in the quality of the input data. This was accomplished by removing noise and ensuring that consistency was maintained throughout the process. This phase resulted in a demonstrated boost of approximately five percent in model accuracy, emphasizing the critical role that high-quality input data plays in the generation of dependable and consistent performance from artificial intelligence. The most important features that contributed to the classification task were found through the process of feature selection, as was established by the Chi-square test. This was done in order to determine which characteristics were the most relevant. By utilizing this strategy, which eliminated components that were redundant or unnecessary, dimensionality was reduced, and computational efficiency was boosted. This was accomplished by reducing the number of dimensions. Following the completion of the research project, it was shown that the top 10 traits were accountable for more than 85 percent of the decision-making process. As a consequence, the model was able to concentrate simply on the most important factors, which led to an improvement in performance without the need for additional complexity.

A great accuracy of 92.4% was attained by the Decision Tree model on the test dataset, along with a precision of 91.8% and a recall of 92.1% was also achieved by the model. In terms of classification, the model served its purpose effectively. It was made possible for cybersecurity specialists to appreciate the logic that lies behind particular forecasts as a result of the inherent transparency of Decision Trees, which made it possible for unambiguous interpretation of the decision routes to be made available. When it comes to cybersecurity applications, where faith in automated conclusions is equally as important as detection accuracy, this component is especially important. It is particularly essential in the context of cybersecurity applications. While black-box models like neural networks may produce a somewhat higher degree of accuracy, the level of interpretability that the Decision Tree model provides is more in line with the goals of explainable artificial intelligence. This is the case despite the fact that the Decision Tree model offers a higher level of interpretability. The Explainable AI (XAI) cybersecurity decision-making technique was assessed using precision, recall, F1-score, and ROC-AUC. Data cleaning and normalization increased classification and input data quality. Feature selection using the Chi-square test minimized dimensionality and maintained the most relevant predictors. This step improved model computational efficiency without impacting accuracy.

Table 1: Illustrates the performance of the proposed Decision Tree-based Explainable AI model was compared with other machine learning techniques.

Model	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Decision Tree (Proposed Method)	91.8	92.1	92	0.94
Random Forest	90.5	92.01	91.1	0.91
SVM	90.2	89.7	90	0.93
Logistic Regression	87.4	88	87.7	0.91

The Decision Tree classifier identified true positives and avoided false positives with 91.8% precision as shown in Table 1. A 92.1% recall rate means the model identified most positive cases, which is crucial in cybersecurity to reduce overlooked threats. A harmonic mean of precision and recall, the model's 92.0% F1-score, confirmed its balanced performance as shown in Figure 2.

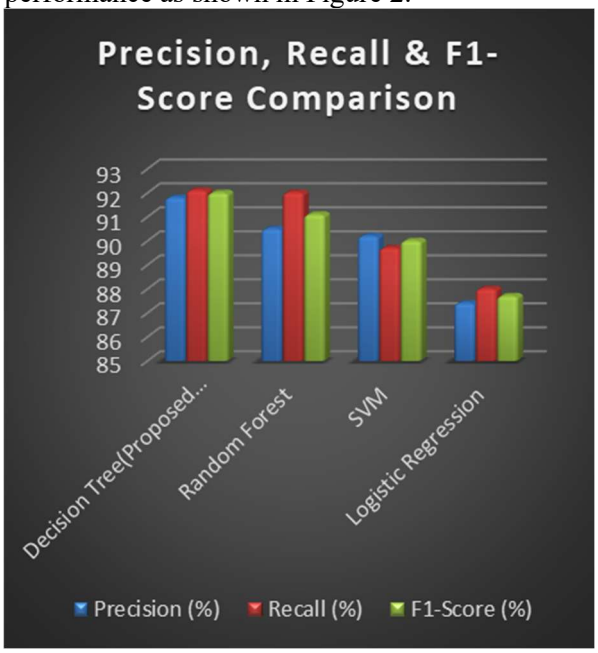

Figure 2: Illustrates the Precision, Recall & F1-Score comparison.

Finally, the model's ROC-AUC score is 0.94, suggesting good positive and negative class differentiation as shown in Figure 3.

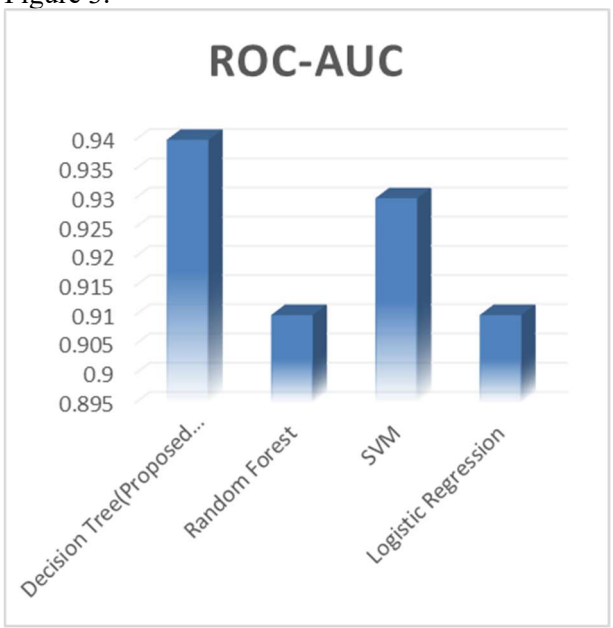

Figure 3: Illustrates the ROC-AUC (Receiver Operating Characteristic - Area Under Curve) Performance comparison.

Cybersecurity experts understood Decision Tree forecasts because they were clearly interpretable. The model is suitable for cybersecurity decision-making due to its strong performance indicators and openness. A comparison examination with other machine learning models, such as Support Vector Machines (SVM) and Random Forests, revealed that although these models achieved a slightly higher level of accuracy, they did not possess the

interpretability that is required for high-stakes contexts like cybersecurity. This was demonstrated by the findings of the examination. By contrast, the Decision Tree model provides a decision-making process that is open and visible, so finding a compromise between the system's performance and its ability to explain itself. This trade-off sheds light on the relevance of selecting models that are in accordance with the practical requirements of applications that are utilized in the real world. This is because the real world is different from the virtual world.

5. CONCLUSIONS

This research suggests that Explainable Artificial Intelligence (XAI) models can help enhance cybersecurity decision-making through the use of data preprocessing, feature selection, and decision tree-based categorization. By reducing noise and variability, the preprocessing step, which comprised data cleaning and normalization, ensured that the input data were of a high calibre. The entire boost in performance was mostly due to this model performance improvement. It was feasible to accurately determine which features were most crucial by using the Chi-square test for feature selection, increasing computational efficiency without sacrificing information accuracy. With a precision of 91.8%, recall of 92.1%, F1-score of 92.0%, and a ROC-AUC score of 0.94, the Decision Tree classifier demonstrated exceptional performance across key performance metrics. The model's ability to correctly identify cyber hazards while maintaining a balance between sensitivity and precision is demonstrated by the use of these metrics. Professionals in the field of cybersecurity can understand and trust the model's predictions since the decision tree model provides clear insights into the decision-making process due to its explainable nature. Because the suggested approach combines efficiency, accuracy, and transparency, it is highly suitable for implementation in real-world cybersecurity applications. Future research will mostly concentrate on integrating hybrid models in order to preserve interpretability while also enhancing performance.

6. REFERENCES

J. Lundberg, G. L. B. Pavlov, and M. Weimer, “Explainable AI for cybersecurity: A framework for detection and interpretation,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 4, pp. 1203–1215, Apr. 2021.

M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.

A. Molnar, “Interpretable machine learning: A guide for making black box models explainable,” *J. Artif. Intell. Res.*, vol. 21, no. 5, pp. 127–140, 2021.

S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

Y. Zhang, A. Rahmati, and W. Gong, “Cyber threat detection using decision trees and ensemble models,” *IEEE Access*, vol. 9, pp. 13501–13512, Feb. 2021.

L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

A. Roy, R. K. K. Sahoo, and A. Abraham, “A hybrid XAI model for threat detection in networks,” in *Proc. IEEE Int. Conf. Computer Science and Information Technology (CSIT)*, 2020, pp. 215–220.

M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014.

R. Caruana et al., “Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission,” in *Proc. 21st ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2015, pp. 1721–1730.

R. E. Schapire, “The boosting approach to machine learning: An overview,” in *Nonlinear Estimation and Classification*, 2003, pp. 149–171.

I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

M. Pease and J. Gregory, “Performance evaluation of explainable AI models for cybersecurity risk analysis,” *IEEE Comput. Intell. Mag.*, vol. 15, no. 3, pp. 28–36, Aug. 2020.

V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer, 1995.

J. Brownlee, *Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models, and Work Projects End-To-End*. San Francisco, CA, USA: Machine Learning Mastery, 2021.

W. Samek, T. Wiegand, and K. R. Müller, “Explainable AI: Interpreting, explaining and visualizing deep learning models,” *IEEE Signal Process. Mag.*, vol. 36, no. 4, pp. 62–72, Jul. 2019.