BOARD #108: DARE-AI: Discovery, Analysis, Research and Exploration Based Experiential Learning Platform for Teaching Integrated AI and Cybersecurity

Utsab Khakurel, Howard University

Utsab Khakurel is a Ph.D. candidate in the Department of Electrical Engineering and Computer Science (EECS) at Howard University, Washington, DC, USA, under the supervision of Dr. Danda B Rawat. Contact him at utsab.khakurel@bison.howard.edu.

Prof. Danda B Rawat, Howard University

Dr. Danda B. Rawat is an Associate Dean for Research & Graduate Studies, a Full Professor in the Department of Electrical Engineering & Computer Science (EECS), Founding Director of the Howard University Data Science & Cybersecurity Center, Founding Director of DoD Center of Excellence in Artificial Intelligence & Machine Learning (CoE-AIML), Director of Cyber-security and Wireless Networking Innovations (CWiNs) Research Lab, and Director of Graduate Cybersecurity Certificate Program at Howard University, Washington, DC, USA. Dr. Danda B. Rawat successfully led and established the Research Institute for Tactical Autonomy (RITA), the 15th University Affiliated Research Center (UARC) of the US Department of Defense as the PI/Founding Executive Director at Howard University, Washington, DC, USA. Dr. Rawat is engaged in research and teaching in the areas of cybersecurity, machine learning, big data analytics and wireless networking for emerging networked systems including cyber-physical systems (eHealth, energy, transportation), Internet-of-Things, multi domain operations, smart cities, software defined systems and vehicular networks. Dr. Rawat has secured over $110 million as a PI and over $18 million as a Co-PI in research funding from the US National Science Foundation (NSF), US Department of Homeland Security (DHS), US National Security Agency (NSA), US Department of Energy, National Nuclear Security Administration (NNSA), National Institute of Health (NIH), US Department of Defense (DoD) and DoD Research Labs, Industry (Microsoft, Intel, VMware, PayPal, Mastercard, Meta, BAE, Raytheon etc.) and private Foundations. Dr. Rawat is the recipient of the US NSF CAREER Award, the US Department of Homeland Security (DHS) Scientific Leadership Award, Presidents' Medal of Achievement Award (2023) at Howard University, Provost's Distinguished Service Award 2021, Researcher Exemplar Award 2019 and Graduate Faculty Exemplar Award 2019 from Howard University, the US Air Force Research Laboratory (AFRL) Summer Faculty Visiting Fellowship 2017, Outstanding Research Faculty Award (Award for Excellence in Scholarly Activity) at GSU in 2015, the Best Paper Awards (IEEE CCNC, IEEE ICII, IEEE DroneCom and BWCA) and Outstanding PhD Researcher Award in 2009. He has delivered over 100 Keynotes and invited speeches at international conferences and workshops. Dr. Rawat has published over 350 scientific/technical articles and 11 books. Dr. Rawat has successfully supervised and graduated 35 PhD students (out of which 28 were under-represented PhD students including 13 female PhD students), successfully supervised 30+ MS students and mentored 7 postdocs, and has been supervising 25 PhD students and mentoring 3 postdocs. Furthermore, he has successfully mentored over 120 minority undergraduate students. He has been serving as an Editor/Guest Editor for over 100 international journals including the Associate Editor of IEEE Transactions on Big Data, Associate Editor of IEEE Transactions on Information Forensics & Security, Associate Editor of Transactions on Cognitive Communications and Networking, Associate Editor of IEEE Transactions of Service Computing, Editor of IEEE Internet of Things Journal, Editor of IEEE Communications Letters, Associate Editor of IEEE Transactions of Network Science and Engineering and Technical Editors of IEEE Network, and Associate Editor of ACM Transactions on Intelligent Systems and Technology. He has been in Organizing Committees for several IEEE flagship conferences such as IEEE INFOCOM, IEEE CNS, IEEE ICC, IEEE GLOBECOM and so on. He served as a technical program committee (TPC) member for several international conferences including IEEE INFOCOM, IEEE GLOBECOM, IEEE CCNC, IEEE GreenCom, IEEE ICC, IEEE WCNC and IEEE VTC conferences. He served as a Vice Chair of the Executive Committee of the IEEE Savannah Section from 2013 to 2017. Dr. Rawat received the Ph.D. degree from Old Dominion University, Norfolk, Virginia. Dr. Rawat served as a Graduate Program Director of Howard Computer Science Graduate Programs (2017 - 2025). Dr. Rawat is a Senior Member of IEEE and a Lifetime Professional Senior

Member of ACM, a Lifetime Member of Association for the Advancement of Artificial Intelligence (AAAI), a Champion Member of USENIX (The Advanced Computing Systems Association), a lifetime member of SPIE, a member of ASEE, a member of AAAS, a member of SAE International, and a Fellow of the Institution of Engineering and Technology (IET). He is an ACM Distinguished Speaker and an IEEE Distinguished Lecturer (FNTC and VTS).

DARE-AI: Discovery, Analysis, Research and Exploration Based Experiential Learning Platform for Teaching Integrated AI and Cybersecurity

Utsab Khakurel and Danda B. Rawat
Department of Electrical Engineering and Computer Science
Howard University, Washington DC, USA
Email: utsab.khakurel@bison.howard.edu, danda.rawat@howard.edu

Abstract

Machine learning (ML) algorithms and artificial intelligence (AI) systems have already had an immense impact on our society. Lately, AI/ML has shown to be able to create machine cognition comparable to or even better than human cognition for some applications. For emerging applications, AI is also regarded to provide cybersecurity solutions (i.e., AI for cybersecurity) by detecting anomalies, adapting security parameters based on ongoing cyberattacks, and reacting in real-time to combat cyber-adversaries. However, ML algorithms and AI systems are vulnerable to manipulation of data or learning models, biases, and low credible information due to flaws in learning models and input data. Therefore, ML algorithms and AI systems need robust security and correctness (i.e., cybersecurity for AI) to permit fair and trustworthy AI systems. Unfortunately, AI and cybersecurity have traditionally been treated as separate fields, with little emphasis on their intersection in education. The primary goal of this paper is to discover, explore, develop and integrate a scalable instructional approach for integrated AI and cybersecurity (DARE-AI for short) in undergraduate and graduate curricula. This is accomplished by creating a "learning by doing" approach to address emerging AI and cybersecurity issues that are not covered in an integrated way, if at all, in traditional curricula. The experiments are designed to study fair and trustworthy AI, adaptive intrusion detection, online learning, federated learning, distributed learning, and adversarial learning. We present learning outcomes and results using surveys and assessments. The developed DARE-AI modules help train the next-generation STEM workforce with knowledge of integrated cybersecurity and AI that is expected to help not only to meet evolving demands of the US government and industries, but also to improve the nation's economic security and preparedness.

1 Introduction

Artificial Intelligence (AI) and Machine Learning (ML) have revolutionized numerous fields, enabling computers to learn from data, recognize patterns, and make autonomous decisions without explicit programming[1,2,3,4]. Unlike traditional rule-based programming, ML algorithms

continuously adapt through data-driven learning without or with minimal human intervention[5]. AI systems and ML algorithms and their various applications such as speech recognition, natural language processing, and computer vision, have been advancing rapidly and demonstrating capabilities comparable to or even surpassing human cognition in certain domains[2,6,7]. One of the most critical applications of AI lies in cybersecurity for emerging applications[8,9,10,11,12,13] by leveraging AI for cyber defense, intrusion detection, anomaly detection, and real-time response to cyber attacks and threats[14,15,16,17,18,19,20]. However, while AI enhances cybersecurity, it is also susceptible to vulnerabilities, including adversarial attacks, biases, and manipulated data[21,22,23,24]. As a result, AI itself requires robust security mechanisms to ensure fairness, reliability, and trustworthiness for cybersecurity for AI. Despite this interdependence, AI and cybersecurity have traditionally been treated as distinct disciplines in education, with little emphasis on their integration. This gap hinders students from gaining a comprehensive understanding of how AI can both enhance and be protected by cybersecurity.

To bridge this divide, we introduce DARE-AI (discover, explore, develop and integrate a scalable instructional approach for integrated AI and cybersecurity) - an experiential learning framework that integrates AI and cybersecurity education through hands-on, research-driven instruction. DARE-AI aims to provide students with a learning-by-doing approach, fostering deep engagement in emerging AI and cybersecurity challenges that are often overlooked in conventional curricula. Through interactive modules, students explore topics such as fair and trustworthy AI, adaptive intrusion detection, online learning, federated learning, distributed learning, and adversarial learning. These experiments, developed using Google Colab with TensorFlow, Pandas, Keras, NumPy, Scikit-learn, Apache Kafka, aif360, SMOTE, SHAP, and more, equip students with practical skills in integrated AI and cybersecurity[25].

Experiential learning (EL) has long been recognized as an effective pedagogical approach in computer science education, fostering active engagement, skill development, and direct interaction with educators[26,27,28]. Research consistently demonstrates that students who engage in EL not only gain a deeper understanding of technical concepts but also develop essential problem-solving, critical-thinking, and communication skills[29,30,27,28]. Furthermore, students in EL-based programs show higher confidence and proficiency in applying their knowledge to real-world scenarios, as evidenced by improved exam performance and technical competence[31]. By integrating EL principles into AI and cybersecurity education, DARE-AI enhances both conceptual learning and practical application, preparing students for the evolving demands of the industry and academia.

The remainder of this paper is organized as follows: Section 2 introduces the DARE-AI research and education program and its objectives. Section 3 provides the overview, learning objectives, dataset and the experiment results of DARE-AI modules for integrated AI and cybersecurity. Section 4 includes the qualitative and quantitative findings of DARE-AI modules. Section 5 briefly discusses the implications of DARE-AI modules and Section 6 presents concluding remarks.

2 DARE-AI: Proposed Research and Education Program

The DARE-AI aims to develop and implement modules, as illustrated in Figure 1, to deliver an integrated research and education program in trustworthy AI and cybersecurity, leveraging cloud-based modules as a service.

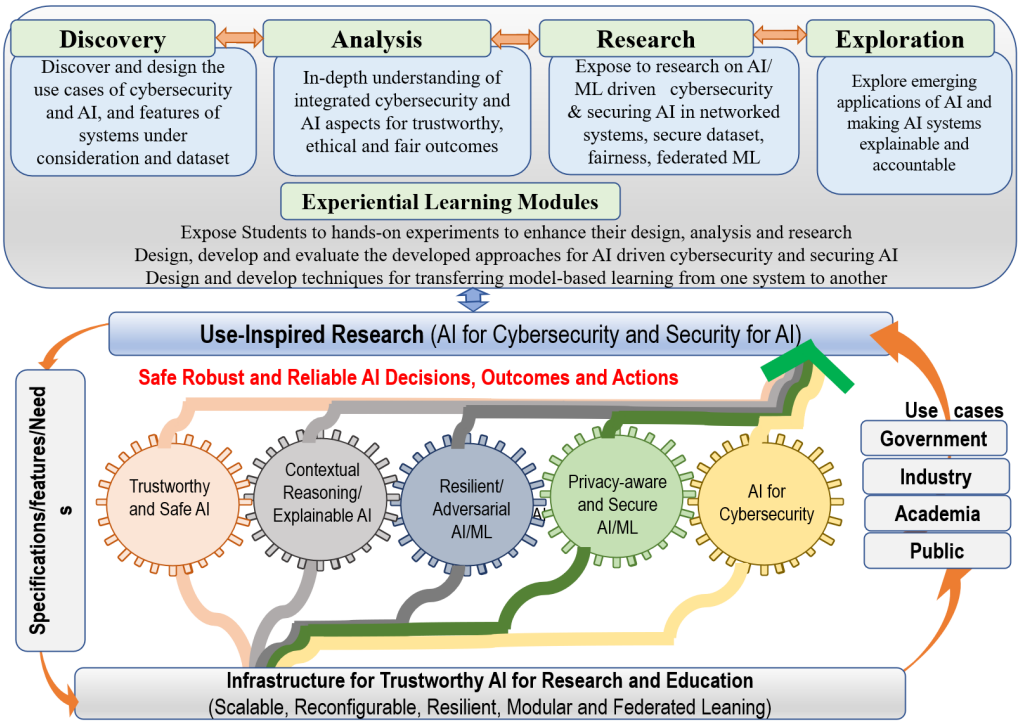

Figure 1: Use-Inspired Research for DARE-AI Hands-on Learning Modules

The DARE-AI modules are structured to provide a comprehensive, hands-on learning experience in AI and cybersecurity: 1) In the 'Discovery' module, students identify and design use cases of cybersecurity and AI discovering applications of AI for Cybersecurity and Cybersecurity for AI, analyze system features, and work with relevant datasets; 2) The 'Analysis' module deepens their understanding of integrated cybersecurity and AI, focusing on the development of trustworthy, safe, and reliable AI/ML algorithms in practical scenarios; 3) In the 'Research' module, students investigate open research questions within emerging areas of machine learning, fostering critical inquiry and innovation; 4) The 'Exploration' module introduces emerging applications, emphasizing AI explainability, accountability, and assurance; and 5) Finally, the 'Experiential' module integrates these concepts through hands-on projects. This modular approach ensures accessibility to students from diverse backgrounds, allowing seamless integration into one or multiple courses.

2.1 Objectives

The primary goal of the interactive laboratories is to provide hands-on experience in key areas of machine learning and cybersecurity while introducing students to emerging research directions.

The core objectives of these experiment-based labs are as follows:

1. Engage students in hands-on experiments to foster interest and active participation in the design, analysis, and research of AI and cybersecurity applications.
2. Develop, implement, and assess strategies for AI-driven cybersecurity and the security of AI systems.
3. Address cybersecurity challenges using AI/ML while strengthening AI/ML algorithms with robust security measures.
4. Explore cutting-edge advancements in ML to drive further research and innovation.

2.2 Discovery Labs

2.2.1 How Machine Learning Learns

Overview: The machine learning phases exhibited in the lab can be explained in simple steps. The input data is gathered, prepared, and split for training and testing. The train data is fed to the chosen machine learning model. This model detects the pattern from the data and the test data is used on the model to assess its competency.

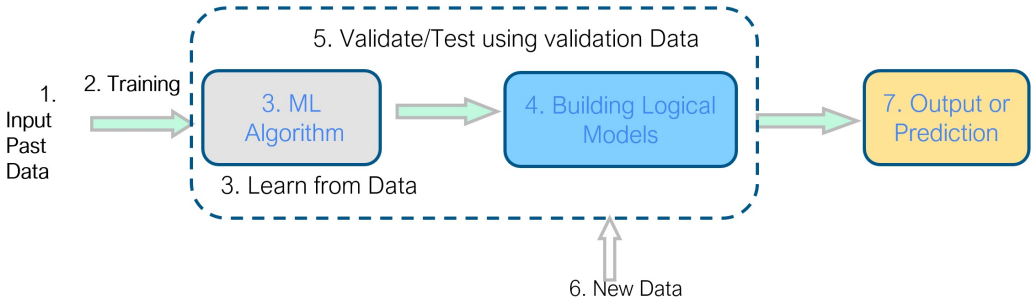

Figure 2: Machine Learning step by step

Learning Objective: The goal of the experiment is to demonstrate how ML model works and to showcase how the model can be trained using input data and used to make predictions on unseen data.

Dataset: The dataset used in this module is devised to demonstrate the workflow of ML models. The data is a bike rental dataset where the feature is temperature of the given day and the target label is the number of bikes rented.

Lab Experiment and Results: The experiment in Lab 1 gives an idea of ML implementation using a simple linear regression model. The linear regression model describes the relationship between the dependent variable, bike rentals, and the independent variable, temperature of the given day. The data is split into a training set and test set. The unknown parameters of the model are calculated using the training data. The parameters slope *(w)* and bias *(b)* obtained from the model are used to visualize the best fit line among the training data points. NumPy polyfit is used

to map the linear relationship represented as $y = mx + c$. Root Mean Square Error (RMSE) is used to estimate the error by measuring the distance between the data points and the regression line. The parameters of the obtained model are *w*=1.46 and *b*=38 respectively. The training RMSE of the model is 31.78. Similarly, test data is used to assess the performance of the model. The validation RMSE of the model is 31.17. Given a data point with independent variable *x*, the error-optimized model can predict a dependent variable *y*. For temperature 53, the predicted bike rental count is 115 compared to the actual value of 114, which is close to actual rentals. Steps with (numbers #) are shown in Figure 3.

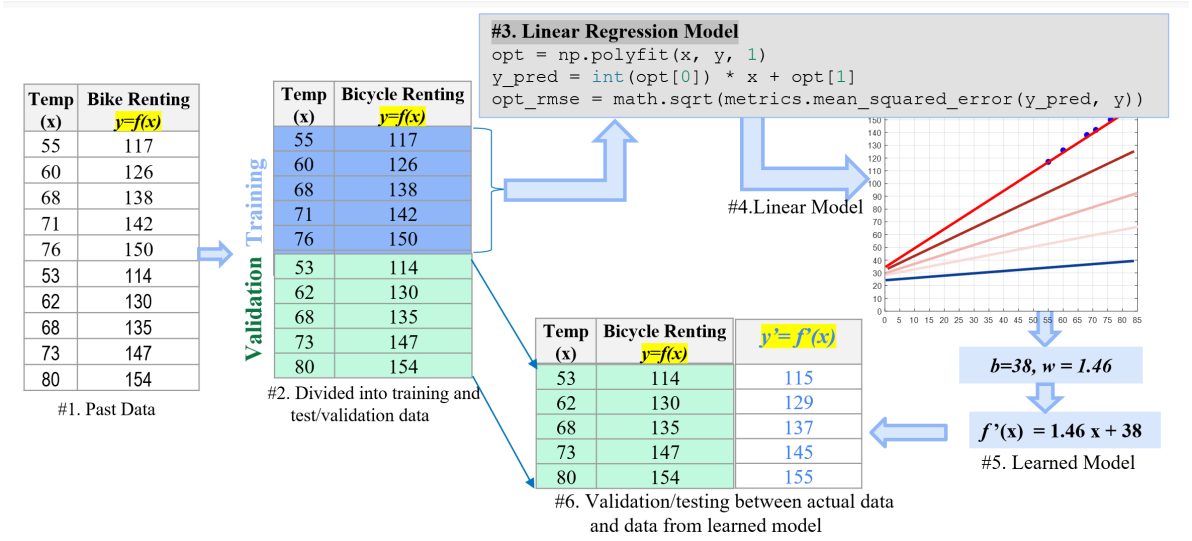

Figure 3: Linear Regression on Bicycle dataset (Actual Rentals vs Predicted Rentals)

2.2.2 Intrusion Detection System for Cybersecurity using Random Forest

Overview: An Intrusion Detection System (IDS) is a security mechanism that monitors network traffic for suspicious activity and potential threats, alerting administrators but not taking direct action[32]. IDS is effective in detecting anomalies and signature-based attacks, but evolving attack techniques often bypass traditional detection methods. This lab explores how machine learning can improve IDS by training models to identify and predict both known and emerging cyber threats, improving overall network security.

Figure 4: Flow diagram for Intrusion Detection System lab

Learning Objective: This module aims to equip students with practical knowledge on using ML algorithms to predict and prevent cyberattacks, enhancing system security before any damage occurs.

Dataset: This lab utilizes an adapted version of the KDD Cup 1999 dataset, a widely recognized benchmark for intrusion detection research. It contains 494,021 instances and 42 attributes, encompassing a diverse range of network attacks[33].

Lab Experiment and Results: The experiment in Lab 2 demonstrates how machine learning can be used to classify cyberattacks before they occur. The KDD dataset is first processed and prepared for training. It includes eight common attack types: smurf, neptune, normal, back, satan, ipsweep, portsweep, and warezclient, listed in descending order of occurrence. However, the dataset is highly imbalanced, with the last five attack types appearing significantly less frequently than the top three. To address this, we apply SMOTE[34], an oversampling technique that balances the dataset by generating synthetic samples for underrepresented attack types.

We use a Random Forest Classifier for this experiment, as it has demonstrated high accuracy and low error rates in intrusion detection[35]. The dataset is split into 80% training, 10% testing, and 10% validation. After training, the model achieves 99% accuracy and 99% precision on the imbalanced dataset. Since rare attack types had minimal representation in the validation set, misclassification rates remained low. Performance metrics improved slightly when using the balanced dataset.

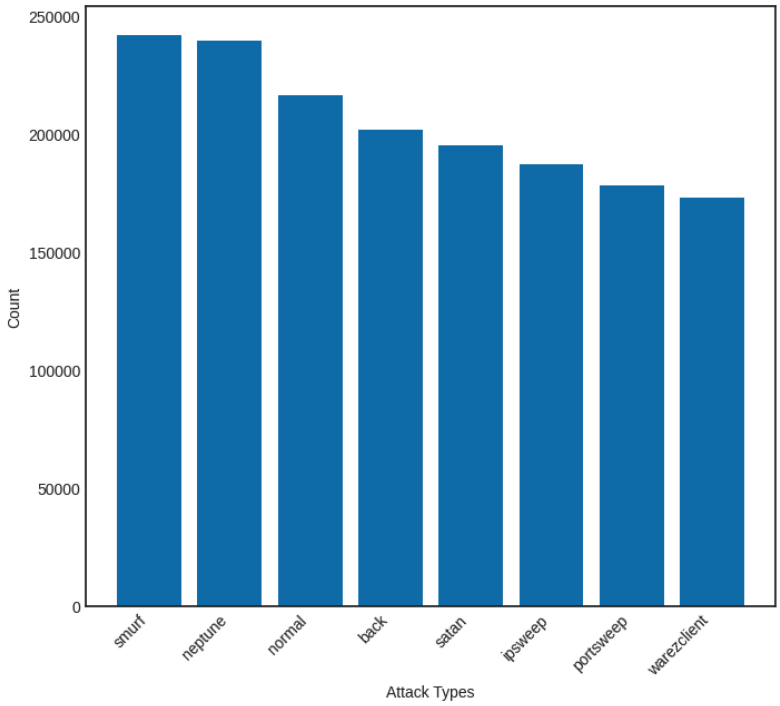

Figure 5: Attack count for top eight attacks in balanced dataset

2.2.3 Detecting and Minimizing Biases in AI

Overview: A bias is an unjustified prejudice in favor of or against a person, group, or thing. Bias in data can lead to unfair and skewed predictions in machine learning models, even when sensitive attributes like gender, race, and ethnicity are removed[36]. This lab explores how biases

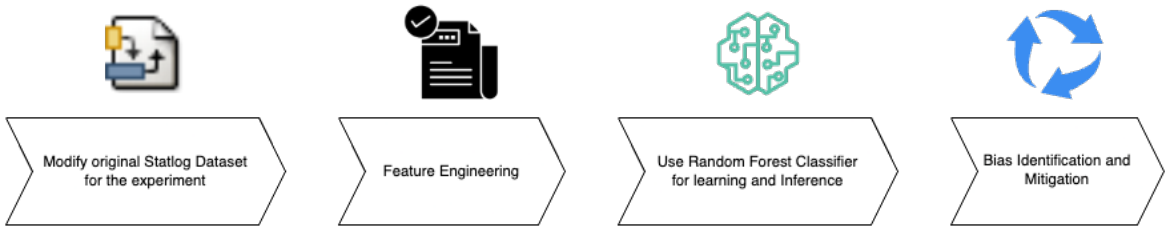

Figure 6: Flow diagram for Detecting and Minimizing biases in AI lab

emerge in datasets and demonstrates methods to identify and mitigate them before training an ML model.

Learning Objective: This module examines bias in data, applies mitigation techniques, and evaluates its impact on the model's classification performance.

Dataset: This lab uses the Statlog Australian Credit Approval dataset from the UCI repository[37], which includes 690 credit card applications with 14 features such as age, employment status, and financial history. The target variable indicates whether a credit application is approved or rejected. The goal is to identify potential gender-based biases in the decision process.

Lab Experiment and Results: The experiment in Lab 3 examines how bias in data affects machine learning models and explores mitigation techniques. The model's performance is evaluated using accuracy, precision, recall, and F1 score. The model's fairness is measured using AIF360 toolkit[38] to identify and mitigate bias in the dataset. Bias is measured using:

- *Statistical Parity Difference (SPD)* is the difference in the rate of favorable outcomes received by the unprivileged group to the privileged group with the ideal fairness value being 0.
- *Disparate Impact (DI)* is the ratio of the rate of a favorable outcome for the unprivileged group to that of the privileged group. The ideal DI value representing fairness is 1.

The protected attribute in this experiment is gender, with males being '1' as the privileged group and females '0' as the unprivileged group. The favorable outcome '1' represents approved applications, while the unfavorable outcome '0' represents rejected applications. For Bias mitigation we use *Reweighing*, which is a preprocessing technique that adjusts instance weights to reduce bias before training. For a privileged group with a favorable outcome, the readjusted weight for that instance is represented as:

$$w = \frac{N_{\text{privileged}} * N_{\text{favorable}}}{N_{\text{all}} * N_{\text{favorable privileged}}} \tag{1}$$

The dataset is split 70% for training and 30% for testing. Random Forest Classifier[39] is used to train models on both biased and bias-mitigated datasets. Statlog Credit Approval dataset has the SPD value of 0.02 and the DI value of 1.05 as presented in Figure 7. The metric reveals a slight bias in favor of females because of the rejection rate for males in training dataset is higher than that for females as shown in Fig 8.

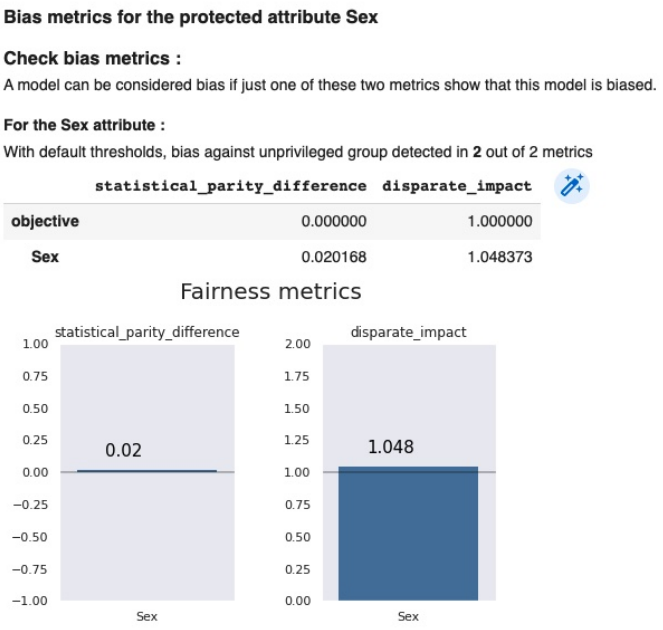

Figure 7: Bias identified in training data through AIF360 bias identification metric tools

Specifically, the Fig. 8 illustrates the distribution of credit card approval and rejection across male and female applicants in the training dataset. While the absolute number of approved male applicants is higher, the approval rate for females (92 out of 198, or 46.5%) slightly exceeds that of males (193 out of 423, or 45.6%). Fairness metrics such as SPD and DI rely on these group-specific proportions rather than raw counts. Since the training data reflects a marginally higher approval rate for females, the model learned to favor this outcome. Consequently, the SPD is 0.02 and DI is 1.05, indicating a slight bias in favor of the unprivileged group in the predicted outcome. This underscores the importance of using rate-based fairness metrics when evaluating model bias, as raw outcome counts alone can be misleading.

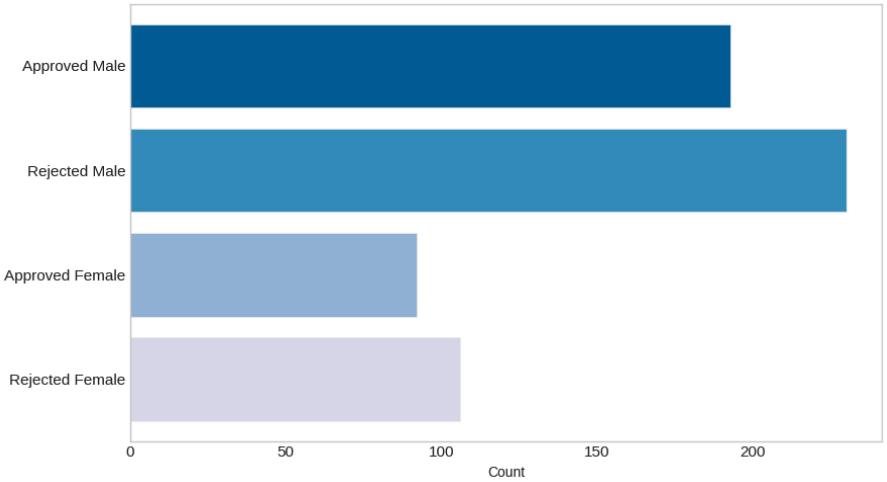

Figure 8: Data distribution of Statlog based on prediction label and sex

```
The accuracy of the model:  0.8653846153846154
The precision of the model:  0.8659994439810954
The recall of the model:  0.8651872399445215
The f1 score of the model:  0.8652725085592672
```

Figure 9: Performance evaluation of the model using biased training data

```
The accuracy of the model:  0.8557692307692307
The precision of the model:  0.855755894590846
The recall of the model:  0.855755894590846
The f1 score of the model:  0.855755894590846
```

Figure 10: Performance evaluation of the model using unbiased training data

Figure 9 presents the classification metrics of the biased model, which achieved 0.865 accuracy, 0.866 precision, and 0.865 recall/F1 score. After reweighing, the model trained on bias-mitigated data had slightly lower performance (0.856 across all metrics) but satisfied the "Error Rate" fairness criterion which requires same False negative and False positive ratio. Figure 10 shows a slight performance drop as bias is reduced, suggesting a trade-off between fairness and model accuracy.

2.3 Analysis Labs

2.3.1 Adversarial Images for Deceiving ML models

Overview: Adversarial images are intentionally modified images that mislead machine learning models while appearing unchanged to the human eye. Attackers exploit this by either corrupting training data or manipulating inputs to force incorrect classifications[40]. This module demonstrates how small pixel modifications can significantly alter model predictions, exposing the vulnerabilities of ML systems to adversarial manipulation.

Learning Objective: This hands-on experiment demonstrates how attackers manipulate images to mislead machine learning models, forcing incorrect classifications.

Dataset: We use MobileNet[41], a pre-trained convolutional neural network (CNN) in Keras, to classify perturbed images. A Chihuahua image from the ImageNet[42] database is extracted, pre-processed, and analyzed to observe the model's response to adversarial modifications.

Lab Experiment and Results: The experiment in Lab 4 demonstrates how adversarial pixel modifications impact MobileNet's classification. The image is perturbed by altering a single

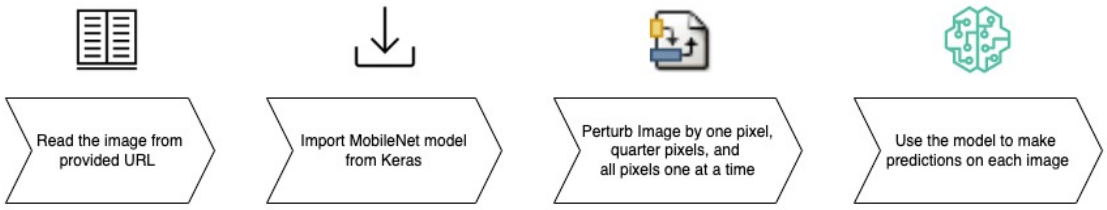

Figure 11: Flow diagram for Adversarial lab

pixel, a quarter of the pixels, and all pixels[43], where the pixels are chosen at random. A single-pixel attack slightly reduces the model's confidence in the correct prediction while keeping the image visually unchanged (Figure 13). As more pixels are altered, the model's accuracy declines, and confidence shifts toward incorrect classifications. In the quarter-pixel attack, the model misclassifies the Chihuahua as a French bulldog with higher confidence (Figure 14). When all pixels are perturbed, the model completely fails to recognize the original class (Figure 15). This experiment highlights how even minor pixel changes can significantly mislead deep learning models.

$$P(x, y) \longrightarrow (x, y) : [r, g, b] + \Delta[r, g, b] \tag{2}$$

```
[[('n02085620', 'Chihuahua', 0.91693276),
  ('n02971356', 'carton', 0.041466217),
  ('n02107312', 'miniature_pinscher', 0.016510503),
  ('n02108915', 'French_bulldog', 0.005804981),
  ('n02096585', 'Boston_bull', 0.0054111257)]]
```

Figure 12: Predictions on an unaltered Chihuahua image

```
[[('n02085620', 'Chihuahua', 0.91371137),
  ('n02971356', 'carton', 0.043758582),
  ('n02107312', 'miniature_pinscher', 0.01696126),
  ('n02108915', 'French_bulldog', 0.0060314927),
  ('n02096585', 'Boston_bull', 0.005356382)]]
```

Figure 13: Predictions on one-pixel perturbed image

```
[[('n02085620', 'Chihuahua', 0.87513745),
  ('n02108915', 'French_bulldog', 0.029640771),
  ('n02971356', 'carton', 0.028842859),
  ('n02107312', 'miniature_pinscher', 0.024846772),
  ('n02096585', 'Boston_bull', 0.0109809125)]]
```

Figure 14: Predictions on quarter-pixel perturbed image

```
[[('n03207941', 'dishwasher', 0.18689789),
  ('n02840245', 'binder', 0.08528389),
  ('n03982430', 'pool_table', 0.0629296),
  ('n04372370', 'switch', 0.0524962),
  ('n04548280', 'wall_clock', 0.03680243)]]
```

Figure 15: Predictions on all pixel perturbed image

2.4 Research Labs

2.4.1 Machine Learning with Event Data Stream for Online Learning

Overview: With the rise of IoT devices, vast amounts of data are continuously generated from various devices/sources[44]. Event streaming allows real-time processing, storage, and analysis of this data to enhance system performance and decision-making[45]. A common example is real-time traffic monitoring, which helps detect collisions and optimize navigation. This module explores online learning using event stream data, demonstrating how Apache Kafka enables real-time model updates.

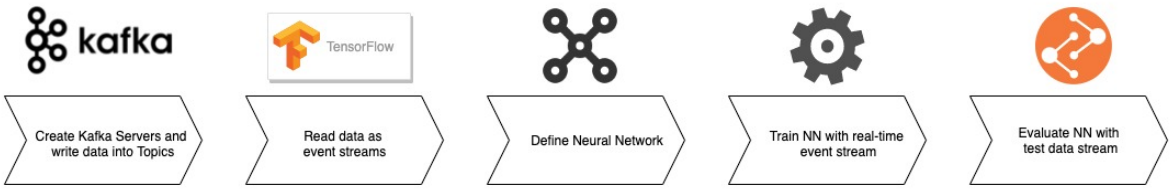

Figure 16: Flow diagram for machine learning with event data stream lab

Learning Objective: This module, adapted from the TensorFlow-IO tutorial[46], demonstrates how to use Apache Kafka for real-time data streaming and online learning. Students will learn to set up and run Kafka and Zookeeper servers, create Kafka topics, and manage data streams by writing records as a producer and reading them as a consumer. The module explores how event stream data is used for training and inference, showcasing the real-time adaptability of machine learning models.

Dataset: The SUSY dataset, generated using Monte Carlo simulation, simulates particle detector data from an accelerator[47]. It contains 5,000,000 instances with 18 attributes and is used for a classification task to predict the presence of supersymmetric particles. The dataset is split 70% for training and 30% for testing. Separate Kafka topics are created for training and testing data, where producers write the training data to susy-train and the test data to susy-test as event streams.

Lab Experiment and Results: The experiment in Lab 5 showcases machine learning with real-time event data streams. Unlike previous modules, this lab uses an Artificial Neural Network (ANN) for training. The model is built using Keras, with Adam optimization[48], accuracy as the performance metric, and binary cross-entropy as the loss function. Real-time data streams are generated using TensorFlow I/O with Kafka and fed into the neural network for training and testing.

```
2188/2188 [==============================] - 30s 13ms/step - loss: 0.4768 - accuracy: 0.7747
<keras.callbacks.History at 0x7f868363b7d0>
```

Figure 17: Loss and Accuracy of the ANN model for train data stream

```
938/938 [==============================] - 23s 24ms/step - loss: 0.4441 - accuracy: 0.7946
```

Figure 18: Loss and Accuracy of the ANN model for test data stream

Figure 17 shows the model's accuracy and loss during training, while Figure 18 presents the test results. The test accuracy closely matches the training accuracy, indicating no overfitting. This experiment highlights how real-time event streams can be effectively used for continuous training and inference in machine learning.

2.4.2 Federated Learning for Data Privacy and Security

Overview: Federated Learning trains machine learning models on decentralized edge devices without transferring data[49,17,18]. Unlike traditional methods, it brings the model to the data, ensuring privacy. Locally trained models are then aggregated to create a fully trained global

model. This module introduces federated learning architecture and its privacy-preserving benefits.

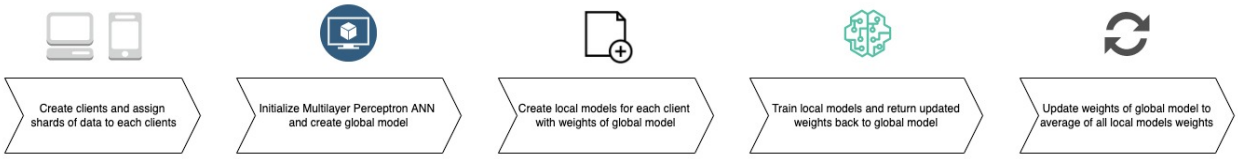

Figure 19: Flow diagram for Federated learning lab

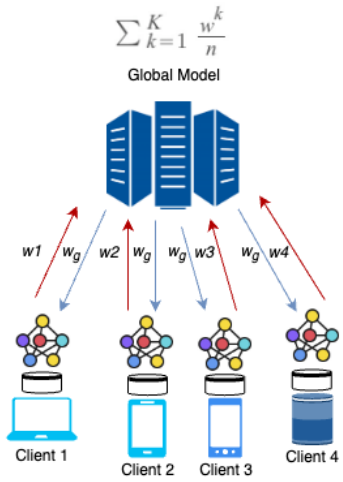

Figure 20: Federated Learning Architecture

Learning Objective: This experiment explores the federated learning architecture and evaluates its performance on an image dataset, demonstrating how decentralized model training operates without data sharing.

Dataset: The CelebA dataset[50], a non-iid dataset, is used for this federated learning experiment. This module utilizes 202,599 celebrity images along with an accompanying CSV file containing facial attributes. The model is trained to predict whether a celebrity is smiling, using the 'smiling' attribute as the classification label.

Lab Experiment and Results: The experiment in Lab 6 illustrates federated learning for privacy-preserving model training[51]. The model is trained using pixel values as features and 'smiling' as the target label from the CelebA dataset. The data is split 90% for training and 10% for testing. A multi-layer perceptron (MLP) is used as the classifier, with stochastic gradient descent (SGD) as the optimizer, accuracy as the performance metric, and binary cross-entropy as the loss function. Due to Colab's memory limitations, an MLP is used instead of a CNN for image classification. The federated learning setup consists of 10 clients, each assigned a data shard for local training. A global model initializes shared weights, which are distributed to all clients. After training on local shards, the updated weights from all clients are aggregated to refine the global model.

Figure 21 presents the accuracy and loss of local and global models. While accuracy is lower than classifiers like Random Forest, this is likely due to limited training data. However, test accuracy

```
9/9 [==============================] - 1s 40ms/step - loss: 0.8416 - accuracy: 0.4403
9/9 [==============================] - 1s 37ms/step - loss: 0.8098 - accuracy: 0.5187
9/9 [==============================] - 1s 38ms/step - loss: 0.8892 - accuracy: 0.4963
9/9 [==============================] - 1s 38ms/step - loss: 0.8905 - accuracy: 0.4701
9/9 [==============================] - 1s 40ms/step - loss: 0.9608 - accuracy: 0.5037
9/9 [==============================] - 1s 39ms/step - loss: 0.7862 - accuracy: 0.5448
9/9 [==============================] - 1s 38ms/step - loss: 0.7647 - accuracy: 0.4925
9/9 [==============================] - 1s 38ms/step - loss: 0.8229 - accuracy: 0.5336
9/9 [==============================] - 1s 38ms/step - loss: 0.8009 - accuracy: 0.5261
9/9 [==============================] - 1s 40ms/step - loss: 0.7229 - accuracy: 0.5224
epoch: 0 | global_acc: 45.82% | global_loss: 0.7998360395431519
9/9 [==============================] - 1s 39ms/step - loss: 0.7583 - accuracy: 0.4403
9/9 [==============================] - 1s 38ms/step - loss: 0.7119 - accuracy: 0.5522
9/9 [==============================] - 1s 39ms/step - loss: 0.7186 - accuracy: 0.5075
9/9 [==============================] - 1s 39ms/step - loss: 0.6965 - accuracy: 0.5485
9/9 [==============================] - 1s 39ms/step - loss: 0.7224 - accuracy: 0.4851
9/9 [==============================] - 1s 39ms/step - loss: 0.7195 - accuracy: 0.4963
9/9 [==============================] - 1s 39ms/step - loss: 0.7119 - accuracy: 0.5672
9/9 [==============================] - 1s 38ms/step - loss: 0.7048 - accuracy: 0.4366
9/9 [==============================] - 1s 37ms/step - loss: 0.7207 - accuracy: 0.4440
9/9 [==============================] - 1s 39ms/step - loss: 0.7194 - accuracy: 0.5336
epoch: 1 | global_acc: 53.18% | global_loss: 0.7156436443328857
```

Figure 21: Accuracy and loss of the local model and global model after each epoch iteration

improves, and test loss decreases after the first iteration, highlighting federated learning's effectiveness.

2.4.3 Distributed Learning for Data Parallelism and Efficiency

Overview: As data volumes exceed the capacity of individual machines, distributed learning enhances processing efficiency, performance, and scalability[52],[53]. This module utilizes data parallelism, where data is evenly distributed across worker nodes, enabling faster training on large datasets. Using Google's Tensor Processing Unit (TPU)[53], the module demonstrates how distributed learning optimizes neural network training.

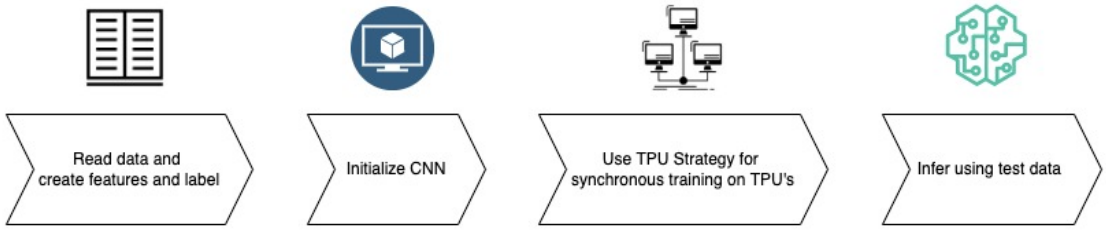

Figure 22: Flow diagram for Distributed learning lab

Learning Objective: This module, similar to federated learning in application and data usage, aims to leverage distributed learning to train a model across multiple nodes and evaluate its performance within a distributed architecture.

Dataset: The CelebA dataset, consisting of 202,599 images, is used in this experiment, similar to the federated learning module[50]. This experiment showcases the use of distributed architectures for efficiently processing large datasets.

Lab Experiment and Results: The experiment in Lab 7 implements model training and inference in a distributed architecture using a large dataset and neural networks. Similar to the federated learning experiment, the model is trained using scaled pixel values as features and the

'smiling' attribute as the target label. To enable distributed learning, the experiment utilizes TPUStrategy, a Google library, leveraging 8 TPU cores in Google Colab. The model follows a synchronous distributed learning approach using all-reduce SGD[54], where each core computes local gradients, and all nodes communicate to aggregate them. The Adam optimizer is used with binary cross-entropy as the loss function.

```
Epoch 1/15
30/30 [==============================] - 30s 494ms/step - loss: 0.8858 - accuracy: 0.5013
Epoch 2/15
30/30 [==============================] - 2s 74ms/step - loss: 0.7069 - accuracy: 0.5054
Epoch 3/15
30/30 [==============================] - 2s 74ms/step - loss: 0.6940 - accuracy: 0.5058
Epoch 4/15
30/30 [==============================] - 2s 74ms/step - loss: 0.6930 - accuracy: 0.5125
Epoch 5/15
30/30 [==============================] - 15s 498ms/step - loss: 0.6924 - accuracy: 0.5147 - val_loss: 0.6902 - val_accuracy: 0.5552
Epoch 6/15
30/30 [==============================] - 2s 73ms/step - loss: 0.6918 - accuracy: 0.5143
Epoch 7/15
30/30 [==============================] - 2s 73ms/step - loss: 0.6909 - accuracy: 0.5132
Epoch 8/15
30/30 [==============================] - 2s 73ms/step - loss: 0.6912 - accuracy: 0.5210
Epoch 9/15
30/30 [==============================] - 2s 75ms/step - loss: 0.6908 - accuracy: 0.5143
Epoch 10/15
30/30 [==============================] - 3s 108ms/step - loss: 0.6896 - accuracy: 0.5147 - val_loss: 0.6898 - val_accuracy: 0.5552
Epoch 11/15
30/30 [==============================] - 2s 74ms/step - loss: 0.6890 - accuracy: 0.5125
Epoch 12/15
30/30 [==============================] - 2s 73ms/step - loss: 0.6870 - accuracy: 0.5143
Epoch 13/15
30/30 [==============================] - 2s 72ms/step - loss: 0.6888 - accuracy: 0.5147
Epoch 14/15
30/30 [==============================] - 2s 73ms/step - loss: 0.6852 - accuracy: 0.5132
Epoch 15/15
30/30 [==============================] - 3s 105ms/step - loss: 0.6820 - accuracy: 0.5173 - val_loss: 0.6894 - val_accuracy: 0.5552
```

Figure 23: Accuracy and loss of the distributed model after each epoch iteration

Unlike the federated learning module, a convolutional neural network (CNN) is used, as distributed learning allows efficient training of deep learning models on large datasets. Validation is performed every 5 epochs. Figure 23 shows that training loss decreases while accuracy improves, with validation loss also declining over epochs, though validation accuracy remains stable. This experiment highlights the scalability and efficiency of distributed learning, enabling neural networks to process large datasets effectively.

2.5 Exploration Labs

2.5.1 Explainable AI

Overview: Explainable AI (XAI) enhances transparency by allowing machine learning models to justify their decisions[55]. It helps interpret model predictions, identify strengths and weaknesses, and anticipate future behavior. This module explores post-hoc explainability using SHAP values to provide insights into the factors influencing the model's decisions.

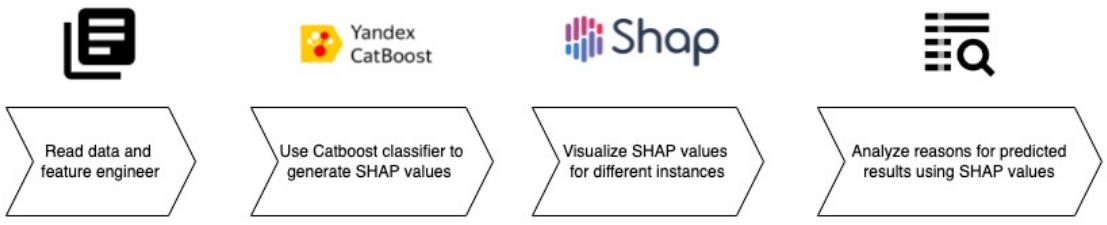

Figure 24: Flow diagram for Explainable AI lab

Learning Objective: This module utilizes SHAP values to assess each feature's contribution to the model's predictions, providing insights into how and why decisions are made based on the data.

Dataset: The Titanic dataset, sourced from the Vanderbilt Department of Biostatistics[56], is used for this experiment. It contains 891 instances with 12 attributes, including passenger details such as gender, age, fare, class, and number of relatives aboard. The target variable, 'survived', indicates whether a passenger survived the disaster. Unique identifiers like Name, Ticket, Cabin, and Passenger ID do not contribute to model learning and are removed from the training set.

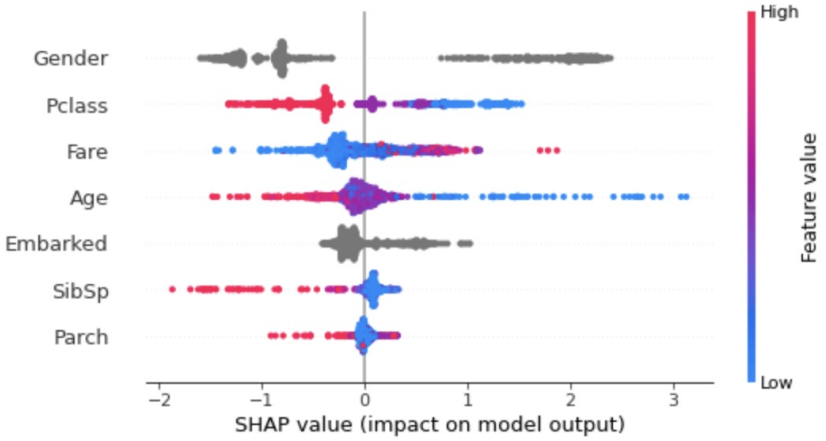

Figure 25: Summary plot for Titanic dataset

Lab Experiment and Results: The experiment in Lab 8 deonstrates the importance of explainable AI (XAI) in building trust in AI systems.The dataset is preprocessed and normalized, and the model is trained using the CatBoost classifier[57], which efficiently handles categorical variables, generating SHAP values for interpretability.

A SHAP summary plot (Figure 25) provides a global view of how features influence predictions. The y-axis ranks features by their impact in descending order, while the x-axis represents SHAP values, indicating whether a feature increases or decreases the probability of survival. Higher feature values are represented in pink, while lower values appear in blue. The plot highlights key trends, showing that gender has a significant impact on survival, even though CatBoost does not provide a direct visual aid for categorical variables. Higher Pclass values, which correspond to lower socioeconomic status, negatively affect survival chances, while higher fare amounts correlate positively with survival, as wealthier passengers had better access to lifeboats.

	Pclass	Gender	Age	SibSp	Parch	Fare	Embarked
SHAP	1.391	2.159	0.265	0.006	0.081	0.702	-0.112

Figure 26: SHAP values representing each feature of the instance 3

To analyze local interpretability, two specific instances from the dataset are examined using SHAP. Instance 3 (Figure 26) represents a female passenger from an upper-class background, where gender had the strongest positive impact on survival, followed by Pclass, Fare, and Age.

	Pclass	Gender	Age	SibSp	Parch	Fare	Embarked
SHAP	0.862	-1.306	-0.046	0.088	0.028	0.53	-0.122

Figure 27: SHAP values representing each feature of the instance 55

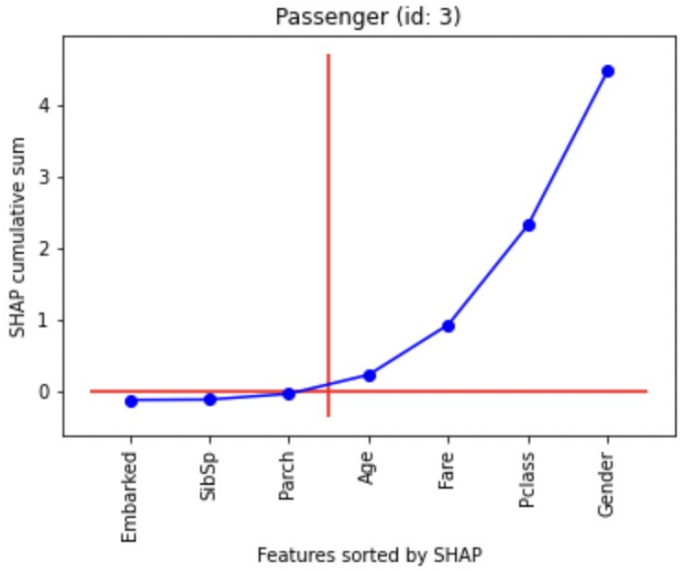

Figure 28: Cumulative SHAP values in ascending order of instance 3

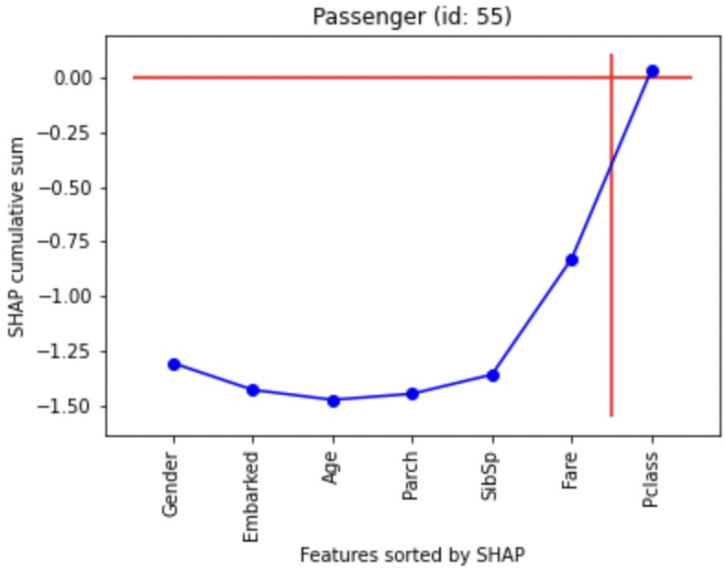

Figure 29: Cumulative SHAP values in ascending order of instance 55

Instance 55 (Figure 27) represents a male passenger who survived despite gender being a negative factor in the model's prediction. His high Pclass and fare ($35.5) played a crucial role in increasing his survival probability. The cumulative SHAP value graphs (Figs. 28, 29) further highlight these trends by selecting the few vital causes from the trivial many, showing that instance 3's survival was driven by gender, wealth, and class, aligning with historical data where

75% of women survived compared to only 19% of men. On the other hand, instance 55 survived solely due to his high social class, with gender contributing the least to his survival.

3 Evaluation and Findings

In this section, we evaluate the effectiveness of the DARE-AI labs through surveys conducted at the end of the module. The analysis of student feedback from the DARE-AI labs, gathered through a structured survey, provided valuable insights into the impact of experiential learning on academic performance, confidence, and conceptual understanding. The evaluation is done both on the quantitative and the qualitative measure. Descriptive statistics were employed to evaluate Likert-scale responses, while categorical data was analyzed to examine student's preferences and engagement with different lab modules.

For the survey, students were sent an evaluation form to collect the feedback on the module. The survey included 20 questions in total, where 11 questions were in the likert scale, 4 were categorical questions and the rest 5 were open ended questions. The specific questions included:

1. Rate the extent to which the DARE-AI labs contributed to improving your understanding of AI/ML and cybersecurity and to improve your grades. (On a scale of 1 to 5 where 1 is no impact and 5 is a significant improvement)
2. Estimate the percentage improvement in your grade after completing the labs. (Select one of 0-10%, 11-20%, 21-30%, 31-40%, 41% or more)
3. How well did the labs prepare you for course assessments (quizzes, exams, assignments)? (On a scale of 1 to 5 where 1 is not at all and 5 is extremely well)
4. How effectively did the labs enhance your understanding of AI/ML concepts (e.g., bias mitigation, adversarial attacks)? (On a scale of 1 to 5 where 1 is not effective and 5 is extremely effective)
5. Rate the improvement in your understanding of cybersecurity principles (e.g., intrusion detection, secure AI) after completing the labs. (On a scale of 1 to 5 where 1 is not effective and 5 is extremely effective)
6. Before completing the labs, how confident were you in understanding integrated AI/ML and cybersecurity concepts? (On a scale of 1 to 5 where 1 is not confident and 5 is extremely confident)
7. After completing the labs, how confident are you in understanding integrated AI/ML and cybersecurity concepts? (On a scale of 1 to 5 where 1 is not confident and 5 is extremely confident)
8. Which lab module had the most impact on improving your grades? (Select one from Discovery Labs, Analysis Labs, Research Labs, Exploration Labs)
9. Rate the usefulness of each lab in helping you better understand course materials: Discovery Labs, Analysis Labs, Research Labs, Exploration Labs. (On a scale of 1 to 5 where 1 is not useful and 5 is very useful)

10. How much did the labs help you develop technical skills directly applicable to course content? (On a scale of 1 to 5 where 1 is no help and 5 is significant help)
11. Rate the following skills in terms of improvement due to the labs: Python, TensorFlow/Keras, Data visualization/analysis, AI/ML tools (aif360, SMOTE, Kafka, SHAP), AI/ML concepts. (On a scale of 1 to 5 where 1 is no improvement and 5 is a significant improvement)
12. How many hours per week did you spend on the DARE-AI labs on average? (Select one from Less than 2 hours, 2-4 hours, 5-7 hours, More than 7 hours)
13. Was the time spent on the labs proportional to the improvement in your grades or understanding? (On a scale of 1 to 5 where 1 is not proportional and 5 is highly proportional)
14. Overall, how satisfied are you with the DARE-AI labs in contributing to your academic success? (On a scale of 1 to 5 where 1 is not satisfied and 5 is highly satisfied)
15. Describe your overall experience with the DARE-AI experiential learning labs. What aspects did you find most engaging?
16. How did the hands-on nature of the labs influence your interest in AI/ML and cybersecurity topics?
17. Did the progression of the labs help you connect theoretical knowledge to practical application? If so, how?
18. Were the instructions and resources provided sufficient to complete the labs? If not, what improvements would you suggest?
19. How did the labs prepare you to use tools like TensorFlow, Python, SHAP, or cybersecurity frameworks in practical scenarios?
20. Would you recommend these labs to other students to help improve their grades or understanding? (Select one from Definitely Yes and Definitely No)

The overall responses to the Likert-scale questions were predominantly positive, reflecting the perceived effectiveness of the DARE-AI labs. For Question 1, which assessed the lab's impact on grade improvement, students reported a mean score of 4.8 and a median of 5.0, indicating a strong consensus that the labs positively influenced academic performance. Similarly, the labs were highly rated for preparing students for assessments (Question 3) and enhancing AI/ML understanding (Question 4), with mean scores of 4.6 and 4.8, respectively, and medians of 5.0, reflecting consistently positive feedback. In terms of cybersecurity understanding (Question 5), students rated the improvement with a mean of 4.5 and a median of 4.5, suggesting that while the labs were effective in this area, perceptions of their impact varied slightly.

Before completing the labs (Question 6), students reported low confidence, with a mean of 2.5 and a median of 2.0, indicating initial uncertainty in the subject matter. After completing the labs (Question 7), confidence significantly improved, with both the mean and median rising to 4.5, demonstrating a strong increase in student's understanding of AI/ML and cybersecurity concepts.A paired t-test was performed to assess whether there was a statistically significant

difference in student's confidence levels before and after completing the labs (Questions 6 and 7). The results showed a statistically significant increase in confidence, with a t-statistic of -5.48 and a p-value of 0.00039. This strongly suggests that the labs had a significant positive effect on student's confidence in their understanding of AI/ML and cybersecurity concepts.

Additionally, when evaluating whether the time spent on the labs was proportional to the improvement in understanding (Question 13), students reported a mean of 4.3, with a median of 4.0, indicating that most students felt the time invested in the labs was appropriately aligned with the gains they made in understanding the material. A correlation analysis between time spent on labs (Question 12) and satisfaction with their academic contribution (Question 14) yielded a coefficient of 0.55, indicating a moderate positive relationship. This suggests that students who invested more time in the labs generally reported higher satisfaction, though the correlation was not particularly strong.

Satisfaction with the lab's contribution to academic success (Question 14) was exceptionally high, with a mean score of 4.9 and a median of 5.0, suggesting a strong consensus among students that the labs significantly contributed to their academic achievement.

The analysis of categorical data provided additional insights into student's experiences with the lab modules (Figure 30). When asked which lab module had the most impact on their grades (Question 8), Discovery Labs were identified as the most impactful by 40% of respondents, followed by Research Labs (30%). The Analysis Labs and Exploration Labs were less frequently chosen, with 20% and 10% of students selecting them as the most impactful, respectively.

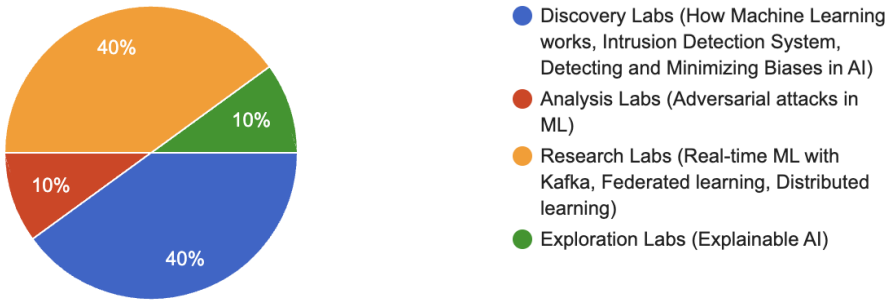

Figure 30: The impact of the DARE-AI modules on student outcomes

In terms of time spent on the labs, most students (60%) spent 5-7 hours per week on the labs, while 30% spent 2-4 hours, and 10% less than 2 hours (Figure 31). These results suggest that students dedicated a significant amount of time to the labs, demonstrating strong engagement with the course material. Furthermore, when asked if they would recommend the labs to other students (Question 20), 100% of respondents answered affirmatively, indicating strong endorsement of the experiential learning approach.

Student ratings on the usefulness of lab modules (Question 9) indicate that Discovery Labs were the most valued, with 9 respondents giving them the highest rating (Figure 32). Research and Analysis Labs also received high ratings, reflecting their strong contribution to learning. While Exploration Labs were rated slightly lower, they still received positive feedback, confirming that

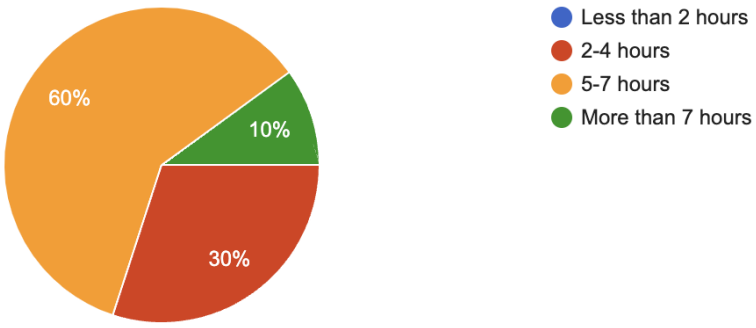

Figure 31: Average weekly time spent on DARE-AI modules

all labs were beneficial, with Discovery Labs having the greatest impact.

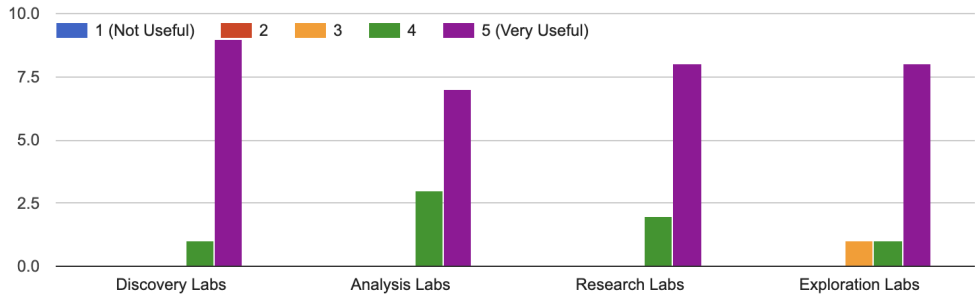

Figure 32: Student feedback on the usefulness of DARE-AI modules

The open-ended survey responses highlight several key themes regarding the DARE-AI experiential learning labs, particularly the hands-on approach, mentorship, and the integration of theory with practical application.

Engagement and Learning Experience: Students found the hands-on nature of the labs highly engaging, particularly the coding exercises and real-time result visualizations. The combination of 1-1 mentorship, project implementation, and exposure to diverse AI/ML concepts kept students motivated and engaged. One student summarized, "The mix of project implementation and mentorship made the labs interesting and valuable."

Impact on Interest in AI/ML and Cybersecurity: The practical, real-world projects significantly increased student's interest in AI/ML and cybersecurity. By actively applying machine learning models, visualizing results, and exploring advanced topics like federated learning and Explainable AI, students gained a deeper curiosity and confidence in these fields. As one student noted, "Working on real projects made the topics more practical and relevant."

Connecting Theory to Practice: The labs were particularly effective in bridging theoretical knowledge with practical application. Students appreciated the step-by-step approach, which allowed them to first understand the theory and then apply it through coding and visualizing results. One respondent highlighted, "The labs connected theory to practice, such as applying bias mitigation concepts using AIF360."

Instruction and Resources: The resources and instructions were generally sufficient, though some students noted that the pace of the labs was fast, requiring additional effort to fully understand the material. Suggestions included providing more examples and troubleshooting guides, especially for advanced topics. However, most students found the clear step-by-step guidance helpful in navigating complex tasks.

Use of Tools and Practical Scenarios: Students expressed high satisfaction with their ability to use industry-standard tools like TensorFlow, Python, and SHAP in practical scenarios. The labs provided hands-on experience in using these tools for real-world applications, with one student noting, "The labs helped me become comfortable with tools like TensorFlow and SHAP, making them relevant to real-world challenges."

The findings strongly support the positive impact of DARE-AI labs on academic performance, confidence, and understanding of AI/ML and cybersecurity. Most students reported grade improvements, with the labs playing a key role in boosting confidence and conceptual grasp. The positive correlation between time spent and satisfaction suggests that greater engagement led to greater benefits. Overall, the high levels of satisfaction, coupled with the unanimous recommendation from students, demonstrate the effectiveness of experiential learning in fostering academic success and deeper engagement with course material.

4 Discussion

The DARE-AI labs effectively enhanced students' academic performance, confidence, and understanding of AI/ML and cybersecurity, with high satisfaction ratings confirming their impact. The positive correlation between time spent and perceived benefits highlights the value of hands-on engagement. Future enhancements could include expanding bias mitigation strategies in the Bias lab, exploring alternative ML models in the IDS lab, and utilizing PyTorch and Scikit-multiflow to deepen insights into online learning. The Adversarial lab could be extended to focus on attack detection and defense, while federated and distributed learning simulations could be refined to reflect real-world implementations. Additionally, integrating explainable AI (XAI) across labs would improve model transparency and interpretability. Overall, the labs successfully bridge theory and practice, and future improvements can strengthen their impact by refining methodologies and expanding research opportunities.

5 Conclusion

Machine learning is driving cutting-edge innovations with a profound impact on our daily life. While AI/ML advancements are significant, security considerations often remain overlooked. Ensuring the protection of systems and devices through advanced ML techniques is crucial for detecting and preventing potential cyber threats. The DARE-AI labs, designed within an experiential learning framework, bridge the gap between AI/ML and cybersecurity. These labs illustrate how ML can enhance security while also addressing vulnerabilities within AI systems. By integrating both practical and theoretical components, they equip students with critical insights into the intersection of AI and cybersecurity. The results of this study demonstrate the effectiveness of the DARE-AI labs in enhancing students' academic performance, boosting their confidence, and deepening their understanding of AI/ML and cybersecurity concepts. Survey

responses indicated that the labs significantly contributed to students' understanding of the topics which helped boost their grades, with a majority reporting improvements in their academic performance. Additionally, students found the hands-on nature of the labs, coupled with 1-1 mentorship, to be highly engaging and beneficial in applying theoretical concepts to real-world scenarios. The labs were also instrumental in enhancing students' interest in AI/ML and cybersecurity topics, providing a practical approach that reinforced the learning process.

In conclusion, the DARE-AI labs successfully contribute to rekindling student interest in data science and security. The combination of interactive learning, practical projects, and real-world applications ensures that students not only grasp essential concepts but are also prepared to tackle real-world challenges. These labs play a crucial role in preparing students for future careers in AI, ML, and cybersecurity, emphasizing the importance of securing AI-driven systems.

Acknowledgments

This research was funded by the DoD Center of Excellence in AI and Machine Learning (CoE-AIML) at Howard University under Contract Number W911NF-20-2-0277 with the U.S. Army Research Laboratory and in part by the US NSF grant CNS/SaTC 2039583. However, any opinion, finding, and conclusions or recommendations expressed in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agencies.

References

[1] K. Das and R. N. Behera, "A survey on machine learning: Concept,algorithms and applications," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 5, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:44614127

[2] S. Kumar and I. Chong, "Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states," *Int. J. Environ. Res. Public Health*, vol. 15, no. 12, p. 2907, Dec. 2018.

[3] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.aaa8415

[4] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Dev.*, vol. 3, pp. 210–229, 1959.

[5] Y. Xu *et al.*, "Artificial intelligence: A powerful paradigm for scientific research," *The Innovation*, vol. 2, no. 4, p. 100179, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666675821001041

[6] G. Yashasree, D. Ganesh, M. Pavan, and K. Bindu, *Applications of Artificial Intelligence Techniques for Cognitive Networks*, 01 2021, pp. 271–284.

[7] J. Jha, A. K. Vishwakarma, C. N, A. Nithin, A. Sayal, A. Gupta, and R. Kumar, "Artificial intelligence and applications," in *2023 1st International Conference on Intelligent Computing and Research Trends (ICRT)*, 2023, pp. 1–4.

[8] B. A. Salau, A. Rawal, and D. B. Rawat, "Recent advances in artificial intelligence for wireless internet of things and cyber–physical systems: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 12 916–12 930, 2022.

[9] D. B. Rawat and C. Bajracharya, "Vehicular Cyber Physical Systems: Adaptive Connectivity and Security," *Springer*, vol. 10, pp. 978–3, 2017.

[10] G. Abdelmoumin, D. B. Rawat, and A. Rahman, "On the performance of machine learning models for anomaly-based intelligent intrusion detection systems for the internet of things," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4280–4290, 2021.

[11] D. B. Rawat, D. C. Popescu, G. Yan, and S. Olariu, "Enhancing VANET performance by joint adaptation of transmission power and contention window size," *IEEE Transactions on parallel and distributed systems*, vol. 22, no. 9, pp. 1528–1535, 2011.

[12] D. B. Rawat and C. Bajracharya, "Detection of false data injection attacks in smart grid communication systems," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1652–1656, 2015.

[13] O. Malomo, D. B. Rawat, and M. Garuba, "A federated cloud computing framework for adaptive cyber defense and distributed computing," in *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2017, pp. 1–6.

[14] M. M. Yamin, M. Ullah, H. Ullah, and B. Katt, "Weaponized ai for cyber attacks," *Journal of Information Security and Applications*, vol. 57, p. 102722, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214212620308620

[15] M. J. Goswami, "Ai-based anomaly detection for real-time cybersecurity," *International Journal of Research and Review Techniques*, vol. 3, no. 1, p. 45–53, Feb. 2024.

[16] D. B. Rawat and K. Z. Ghafoor, *Smart cities cybersecurity and privacy*. Elsevier, 2018.

[17] B. Ghimire and D. B. Rawat, "Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for internet of things," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8229–8249, 2022.

[18] A. Uprety and D. B. Rawat, "Reinforcement learning for IoT security: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 11, pp. 8693–8706, 2020.

[19] D. B. Rawat, R. Doku, and M. Garuba, "Cybersecurity in big data era: From securing big data to data-driven security," *IEEE Transactions on Services Computing*, vol. 14, no. 6, pp. 2055–2072, 2019.

[20] P. K. Singh, S. Nandi, S. K. Nandi, U. Ghosh, and D. B. Rawat, "Blockchain meets ai for resilient and intelligent internet of vehicles," *arXiv preprint arXiv:2112.14078*, 2021.

[21] K. DeMedeiros, A. Hendawi, and M. Alvarez, "A survey of ai-based anomaly detection in iot and sensor networks," *Sensors*, vol. 23, no. 3, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/3/1352

[22] A. Rawal, J. McCoy, D. B. Rawat, B. M. Sadler, and R. S. Amant, "Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 6, pp. 852–866, 2021.

[23] S. Silva and M. Kenney, "Algorithms, platforms, and ethnic bias: An integrative essay," *Phylon (1960-)*, vol. 55, no. 1 2, pp. 9–37, 20x18. [Online]. Available: https://www.jstor.org/stable/26545017

[24] F. O. Olowononi, D. B. Rawat, and C. Liu, "Resilient machine learning for networked cyber physical systems: A survey for machine learning security to securing machine learning for cps," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 524–552, 2020.

[25] D. Rawat and U. Khakurel, "Machine learning @ hu," 2021. [Online]. Available: https://sites.google.com/view/intro2ml-fall2021/labs?authuser=0

[26] Z. Kissel and C. Stuetzle, "Experiential learning framework for smaller computer science programs," 02 2020.

[27] D. B. Rawat and C. Bajracharya, "Informed teaching and learning using thought-bubbles for focusing student attention in engineering courses," in *2016 ASEE Annual Conference & Exposition*, 2016.

[28] ——, "Enhancing student learning through proactive feedback based adaptive teaching for engineering courses," *International Journal on Integrating Technology in Education*, vol. 4, no. 3, 2015.

[29] P. Sendall, C. Stuetzle, Z. Kissel, and T. Hameed, "Experiential learning in the technology disciplines," 02 2020.

[30] N. Caporusso, "An experiential learning approach to research methods in computer science based on smart goals," in *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 2020, pp. 802–807.

[31] R. Borela, Z. Liding, and M. McDaniel, "Enhancing cs1 education through experiential learning with robotics projects," in *Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1*, ser. SIGCSETS 2025. New York, NY, USA: Association for Computing Machinery, 2025, p. 144–150. [Online]. Available: https://doi.org/10.1145/3641554.3701810

[32] A. S. Ashoor and S. D. Gore, "Intrusion detection system (ids) &intrusion prevention system (ips): Case study," 2011.

[33] S. Hettich and S. D. Bay, "The uci kdd archive," 1999. [Online]. Available: http://kdd.ics.uci.edu

[34] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, p. 321–357, Jun 2002. [Online]. Available: http://dx.doi.org/10.1613/jair.953

[35] M. Almseidin, M. Alzubi, S. Kovacs, and M. Alkasassbeh, "Evaluation of machine learning algorithms for intrusion detection system," 2018.

[36] S. Corbett-Davies and S. Goel, "The measure and mismeasure of fairness: A critical review of fair machine learning," 2018.

[37] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[38] R. K. E. Bellamy, K. Dey, M. Hind, S. C. Hoffman, S. Houde, K. Kannan, P. Lohia, J. Martino, S. Mehta, A. Mojsilovic, S. Nagar, K. N. Ramamurthy, J. T. Richards, D. Saha, P. Sattigeri, M. Singh, K. R. Varshney, and Y. Zhang, "AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias," *CoRR*, vol. abs/1810.01943, 2018. [Online]. Available: http://arxiv.org/abs/1810.01943

[39] Scikit-learn, "sklearn.ensemble.RandomForestClassifier," 2021.

[40] G. R. Machado, E. Silva, and R. R. Goldschmidt, "Adversarial machine learning in image classification: A survey toward the defender's perspective," *ACM Computing Surveys*, vol. 55, no. 1, p. 1–38, Jan 2023. [Online]. Available: http://dx.doi.org/10.1145/3485133

[41] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017.

[42] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and F.-F. Li, "Imagenet: a large-scale hierarchical image database," 06 2009, pp. 248–255.

[43] J. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," *IEEE Transactions on Evolutionary Computation*, vol. 23, no. 5, p. 828–841, Oct 2019. [Online]. Available: http://dx.doi.org/10.1109/TEVC.2019.2890858

[44] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for iot big data and streaming analytics: A survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 4, pp. 2923–2960, 2018.

[45] D. Kifer, S. Ben-David, and J. Gehrke, "Detecting change in data streams," 04 2004, pp. 180–191.

[46] T. IO, "Robust machine learning on streaming data using kafka and tensorflow-io," https://github.com/tensorflow/io/blob/master/docs/tutorials/kafka.ipynb, 2022.

[47] P. Baldi, P. Sadowski, and D. Whiteson, "Searching for Exotic Particles in High-Energy Physics with Deep Learning," *Nature Commun.*, vol. 5, p. 4308, 2014.

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[49] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *Foundations and trends® in machine learning*, vol. 14, no. 1–2, pp. 1–210, 2021.

[50] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[51] SaTC-AI. (2021) Dare-ai modules. [Online]. Available: https://github.com/DARE-AI/DARE-AI

[52] A. Galakatos, A. Crotty, and T. Kraska, *Distributed Machine Learning*, 01 2017, pp. 1–6.

[53] J. Verbraeken, M. Wolting, J. Katzy, J. Kloppenburg, T. Verbelen, and J. S. Rellermeyer, "A survey on distributed machine learning," *ACM Comput. Surv.*, vol. 53, no. 2, mar 2020. [Online]. Available: https://doi.org/10.1145/3377454

[54] P. H. Jin, Q. Yuan, F. Iandola, and K. Keutzer, "How to scale distributed deep learning?" 2016.

[55] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, 2021. [Online]. Available: https://www.mdpi.com/1099-4300/23/1/18

[56] S. Molony. (2011) Titanic : Triumph and tragedy. [Online]. Available: https://www.encyclopedia-titanica.org/titanic-triumph-and-tragedy-eaton-and-haas.html

[57] A. V. Dorogush, V. Ershov, and A. Gulin, "Catboost: gradient boosting with categorical features support," 2018.