Authentic Learning on Machine Learning for Cybersecurity

Dan Chia-Tien Lo
Department of Computer Science
Kennesaw State University
Marietta, Georgia, USA
dlo2@kennesaw.edu

Hossain Shahriar
Department of Information Technology
Kennesaw State University
Marietta, Georgia, USA
hshahria@kennesaw.edu

Kai Qian
Department of Computer Science
Kennesaw State University
Marietta, Georgia, USA
kqian@kennesaw.edu

Michael Whitman
Institute for Cybersecurity Workforce Development
Kennesaw State University
Marietta, Georgia, USA
mwhitman@kennesaw.edu

Fan Wu
Department of Computer Science
Tuskegee University
Tuskegee, Alabama, USA
fwu@tuskegee.edu

Cassandra Thomas
Department of Computer Science
Tuskegee University
Tuskegee, Alabama, USA
cthomas@tuskegee.edu

ABSTRACT

The primary goal of the authentic learning approach is to engage and motivate students in learning real world problem solving. We report our experience in developing k-nearest neighbor (KNN) classification for anomaly user behavior detection, one of the authentic machine learning for cybersecurity (ML4CybrS) learning modules based on 10 cybersecurity (CybrS) cases with machine learning (ML) solutions. All portable labs are made available on Google CoLab. So students can access and practice these hands-on labs anywhere and anytime without software installation and configuration which will engage students in learning concepts immediately and getting more experience for hands-on problem solving skills.

1 AUTHENTIC LABWARE DESIGN

The need for virtual labs have been a great demand over last decades and the Covid-19 pandemic had pushed virtual learning to the limit. Among many benefits, the key advantages of virtual labs are self-paced personalized learning at anytime and anywhere without the costly lab space and facilities, which tends to alleviate learning inequality with students from socioeconomically disadvantaged households. As a guide in designing such labware, we aim to employ common affordable hardware with open source software that may be replicated at other institutions, especially those with resource deficit. Our authentic portable labware is developed, and deployed on Google CoLab where learners can access, code, share, and practice all labs interactively with only web browsers anywhere and anytime with zero installation and configuration. Each learning module is designed based on a specific real world cybersecurity case with a step-by-step instruction to process a real-world data set.

SIGCSE 2023, March 15–18, 2023, Toronto, ON, Canada

ACM ISBN 978-1-4503-9433-8/23/03.
https://doi.org/10.1145/3545947.3576245

2 LEARNING ASSESSMENT AND ANALYSIS

The M10 learning module, KNN classification for anomaly user behavior detection, was used for a homework assignment in CS 4265, Spring, 2022, for 39 students. Though the class is face to face, we try to give them the learning module without any other information, as if they were online self-learning students. The students are exposed to the pre-lab, which contains the prior knowledge of the KNN algorithm and anomaly user behaviors. A sample KNN program developed in Google CoLab is given to illustrate the pros and cons of KNN. Followed by the pre-lab is the in-lab, which dives into detailed KNN programming with a short video that explains the dataset, the algorithm, and the result. After the student complete pre-lab and in-lab, they are asked to complete a post-lab assignment and turned in their work. We ask them to extend the in-lab program to find a better value of k in terms of the classification accuracy. The search should start from half of k to two times of k, i.e., $\frac{k}{2} \sim 2k$. The k value with the highest accuracy will be printed out.

As a result, 35 out of 39 students complete the homework with a score of 90% and above. 4 students do not submit their work, whose final grades are C or D. The result is very encouraging that all students are able complete the assignment by self-learning the materials. Another interesting finding is students' reflection on learning cybersecurity and machine learning. 2.30/5.0 think they are receiving CybrS education, 3.45/5.0 for ML education, and 1.64/5.0 for ML4CybrS. The result shows our CS program may need to add extra cybersecurity courses and definitely need to teach more ML4CybrS courses. When asked about hands-on labware, students rate highest for "learn better by personally doing or working through examples," (4.64/5.0). Only 2.77/5.0 is rated for "learn better by listening to lectures," and 3.16/5.0 for "learn better by reading the materials on my own." Overall, 95.12% of students like to work with Google CoLab, the learning module helps them learn ML4CybrS.

3 CONCLUSION

CybrS and ML are both important topics in CS curriculum but the combination of the two could be a future trend in this field and thus the need to develop labware is obvious. Our preliminary results indicate that our labware design is effective in both learning and cost.