Cybersecurity Workforce Upskilling: CRISP-DM Automation with Jupyter Notebooks and Immersive Lab Terraform Modules

Article *in* International Journal of Science and Research (IJSR) · April 2025

DOI: 10.21275/SR25408010102

1 author:

Sandhya Guduru
HCL America
20 PUBLICATIONS **6** CITATIONS

SEE PROFILE

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

Cybersecurity Workforce Upskilling: CRISP-DM Automation with Jupyter Notebooks and Immersive Lab Terraform Modules

Sandhya Guduru

Independent Researcher, Masters in Information Systems Security

Abstract: ***The rapid evolution of cybersecurity threats necessitates continuous workforce upskilling, with automation playing a crucial role in enhancing training effectiveness. This research explores the integration of the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework with Jupyter Notebooks for automating threat modeling and security analysis. Additionally, it examines the use of Immersive Lab Terraform modules to provide hands-on cloud attack simulations, bridging the gap between theoretical knowledge and practical application. By aligning these methodologies with CyberSeek Heatmap metrics, this approach ensures targeted skill development that meets industry demands. The study highlights the potential of automation-driven training pipelines in equipping cybersecurity professionals with the necessary expertise to mitigate evolving threats.***

Keywords: Cybersecurity Workforce Upskilling, CRISP-DM, Jupyter Notebooks, Threat Modeling, Immersive Lab Terraform Modules, Cloud Security, CyberSeek Heatmap, Security Automation, Hands-on Training, Skill Gap Analysis

1. Introduction

As cyber threats evolve in complexity, the need for an advanced, well-trained cybersecurity workforce has become increasingly critical. Traditional training approaches often emphasize theoretical knowledge without providing sufficient hands-on experience, limiting professionals' ability to respond effectively to real-world security challenges. To address this gap, automation-driven learning frameworks are incorporated into cybersecurity training to enhance practical skill development.

One such framework is the Cross-Industry Standard Process for Data Mining (CRISP-DM), established in 1996 as a structured methodology for data-driven projects. It consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. While initially designed for data mining applications, its structured approach provides a robust foundation for cybersecurity training by organizing learning into clear, actionable steps [1].

Integrating CRISP-DM with Jupyter Notebooks enables learners to engage in real-time security analysis, dynamic coding, and interactive visualization, fostering a more practical learning experience. Additionally, Immersive Labs' Terraform modules offer cloud-based simulations of cyberattacks, allowing professionals to practice responding to security incidents in controlled environments. These modules cover key areas such as threat detection, digital forensics, and incident response.

To ensure training remains aligned with industry demands, cybersecurity education programs can leverage the CyberSeek Heatmap, which provides insights into current job market trends and workforce skill gaps. This research explores the integration of CRISP-DM, Jupyter Notebooks, and Immersive Labs' Terraform modules as a structured approach to cybersecurity workforce upskilling. By combining these methodologies, training programs can better equip professionals with the technical expertise needed to address modern cybersecurity threats effectively.

2. Literature Review

In the rapidly evolving field of cybersecurity, continuous upskilling of the workforce is essential to counter emerging threats. Integrating methodologies like the Cross-Industry Standard Process for Data Mining (CRISP-DM) with tools such as Jupyter Notebooks and Immersive Lab Terraform modules offers a structured approach to enhance cybersecurity training programs.

CRISP-DM provides a comprehensive framework for data mining projects, comprising six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. This methodology has been widely adopted across industries for its flexibility and robustness in tackling data-driven challenges [2].

Jupyter Notebooks have emerged as a powerful tool in cybersecurity analysis, enabling professionals to conduct threat hunting, automate repetitive tasks, and analyze extensive datasets. By combining live code with explanatory text, Jupyter facilitates documentation and sharing of methodologies, enhancing collaborative efforts in threat detection and response [3].

The integration of CRISP-DM with Jupyter Notebooks in cybersecurity training allows for the automation of threat modeling processes. For instance, the Jupyter-Threat project demonstrates how to initiate threat modeling sessions within the software development lifecycle, utilizing Jupyter to document and analyze threats systematically [4].

Furthermore, Immersive Labs offers cloud security challenges that require teams to secure cloud environments through real-world attack scenarios and misconfiguration challenges. These hands-on labs and simulations improve detection, response, and remediation skills across various

cloud platforms, ensuring teams are prepared to defend against evolving cloud threats [5].

Aligning these training initiatives with frameworks like the CyberSeek Heatmap ensures that skill development efforts are targeted towards identified workforce gaps. By mapping training outcomes to specific skill shortages, organizations can strategically address areas of need, enhancing the overall competency of their cybersecurity teams. Implementing skill mapping tools allows organizations to assess existing employee capabilities against required competencies, effectively identifying and addressing skill gaps. Additionally, frameworks like the Cybersecurity Body of Knowledge (CyBOK) can be utilized to develop comprehensive knowledge profiles of the workforce, facilitating targeted training initiatives. By aligning training programs with these identified gaps, organizations ensure that their cybersecurity teams are equipped with the necessary skills to combat evolving threats [6].

Training Method	Advantages	Limitations
Traditional Online Courses	Theoretical knowledge, flexible learning.	Limited hands-on experience.
Certification Programs	Industry recognition, structured curriculum.	Expensive, often lacks real-world scenarios.
Jupyter Notebooks for Training	Interactive coding, real-time threat analysis.	Requires programming knowledge.
Immersive Lab Terraform Modules	Hands-on cloud attack simulations.	Needs cloud infrastructure access.

Figure 1: Comparison of Cybersecurity Training Methods

This literature review underscores the value of integrating structured methodologies like CRISP-DM with interactive tools such as Jupyter Notebooks and Immersive Lab Terraform modules in cybersecurity workforce upskilling. Such integrations facilitate a comprehensive, hands-on approach to training, effectively bridging the gap between theoretical knowledge and practical application in the ever-evolving cybersecurity landscape.

3. Problem Statement

The increasing complexity of cyber threats necessitates continuous upskilling of cybersecurity professionals. Traditional training methods, such as static coursework and theoretical assessments, often fail to provide the hands-on experience required to detect, analyze, and mitigate real-world attacks. As cybercriminal tactics evolve, there is a widening skills gap, with organizations struggling to equip their security teams with the latest threat detection and response capabilities.

A significant limitation in current cybersecurity training programs is the lack of integration between structured methodologies and interactive tools. For instance, the Cross-Industry Standard Process for Data Mining (CRISP-DM) provides a structured approach to data analysis, but its application in cybersecurity training remains limited. Similarly, while Jupyter Notebooks offer an interactive computing environment beneficial for data analysis and modeling, their potential for automating threat modeling and simulation exercises is underutilized. Integrating CRISP-DM with Jupyter Notebooks could enhance the effectiveness of cybersecurity training by providing a structured yet flexible framework for developing and testing threat models [7].

Moreover, the adoption of Infrastructure as Code (IaC) tools like Terraform has revolutionized the management of cloud infrastructure, enabling the automation of deployment and scaling processes. However, the use of Terraform modules for creating immersive lab environments that simulate cloud attack scenarios is not widespread in current training programs. Leveraging Terraform for such simulations can provide cybersecurity professionals with practical experience in identifying and mitigating cloud-specific threats, thereby enhancing their readiness to protect cloud environments.

To address these challenges, there is a need for a comprehensive training framework that combines CRISP-DM, Jupyter Notebooks, and Terraform-based labs. Such an integrated approach would bridge the gap between theoretical knowledge and practical application, ensuring that cybersecurity professionals are equipped with the necessary skills to defend against evolving cyber threats. By aligning training initiatives with frameworks like the CyberSeek Heatmap, organizations can tailor their upskilling efforts to address specific skill shortages, thereby enhancing the overall competency of their cybersecurity workforce.

Challenge	Current Limitation	Proposed Solution
Lack of Hands-on Experience	Traditional training relies on static coursework and assessments, offering minimal real-world exposure.	Integrate interactive labs with practical simulations.
Limited Use of CRISP-DM in Cybersecurity	CRISP-DM is widely used for data analysis but not fully utilized in cybersecurity training.	Apply CRISP-DM to guide structured threat modeling.
Underutilization of Jupyter Notebooks	Jupyter Notebooks provide an interactive environment but are rarely used for threat simulation.	Automate threat modeling and security analysis with Jupyter.
Ineffective Cloud Security Training	Many programs lack Terraform-based cloud attack simulations.	Use Terraform modules to create immersive cloud security labs.
Misalignment with Industry Needs	Training programs do not align with real-time industry skill gaps.	Align training with CyberSeek Heatmap metrics to address workforce shortages.

Figure 2: Gaps in Cybersecurity Training Programs

Current Training Limitations

The cybersecurity industry faces a growing skills gap, with demand for professionals outpacing the availability of qualified talent. Traditional cybersecurity training methods, such as static online courses and theoretical certification programs, often fail to provide the hands-on experience necessary for real-world threat mitigation. Research indicates that non-interactive training approaches are frequently

ineffective in preparing individuals for practical cybersecurity challenges. Additionally, studies have shown that employees often do not exhibit significant improvement in security practices following traditional training sessions, highlighting the need for more engaging and practical training methodologies [8]. Despite the use of structured frameworks like CRISP-DM, these methodologies are rarely integrated into training pipelines, resulting in a lack of data-driven threat modeling expertise among professionals.

Moreover, traditional Jupyter-based security training relies on pre-defined datasets and lacks dynamic simulation capabilities. This approach limits learners' ability to engage with real-time attack scenarios, reducing their preparedness for evolving cyber threats. Terraform-based cloud security training often focuses on infrastructure provisioning rather than offensive and defensive security simulations, making it insufficient for comprehensive workforce upskilling [9].

Furthermore, industry-aligned skill benchmarks such as the CyberSeek Heatmap indicate that many professionals lack experience with automation-driven threat detection and cloud security configurations [10]. Without incorporating automated tools into cybersecurity education, training programs fail to address real-world security challenges effectively.

Emerging Technologies in Cybersecurity Training

Recent advancements in cybersecurity education leverage automation, immersive simulations, and machine learning-driven analytics to enhance workforce upskilling. Platforms such as Jupyter Notebooks have become instrumental in cybersecurity training by integrating AI-based threat detection models, enabling learners to dynamically analyze real-world attack patterns. For instance, the GitHub repository "Applications-of-AI-for-Anomaly-Detection" offers workshops demonstrating the use of AI to detect anomalies in time-series data, which is crucial for identifying potential security threats [11].

Additionally, tools like NVIDIA's Morpheus provide AI-based solutions for real-time cybersecurity threat detection and alert prioritization, which can be integrated within Jupyter Notebooks to enhance training and analysis capabilities [12].

Unlike traditional rule-based approaches, these intelligent systems provide adaptive learning experiences tailored to evolving cybersecurity threats [13].

Terraform-based cloud security training has also evolved with the introduction of Immersive Lab modules, which offer interactive scenarios that simulate complex cyberattacks. These modules enable cybersecurity professionals to practice cloud misconfiguration detection, privilege escalation, and automated remediation techniques in a controlled environment [14]. Additionally, modern training platforms incorporate reinforcement learning models to personalize security exercises, ensuring that professionals acquire skills relevant to their job roles.

Furthermore, CyberSeek Heatmap-aligned training programs integrate real-time industry metrics, bridging the gap between academic knowledge and job market requirements. These solutions provide tailored skill development roadmaps based on current cybersecurity demand trends, improving workforce readiness [15]. The integration of automation-driven methodologies into cybersecurity education is crucial for addressing emerging threats and mitigating skill gaps effectively.

4. Proposed Solution

To address the evolving challenges in cybersecurity, an integrated training framework that combines the Cross-Industry Standard Process for Data Mining (CRISP-DM), Jupyter Notebooks, and Terraform-based immersive labs is proposed. This approach aims to provide structured data analysis, interactive learning environments, and hands-on cloud attack simulations to enhance the skills of cybersecurity professionals.

CRISP-DM for Structured Threat Modeling

To systematically enhance cybersecurity training, the CRISP-DM framework provides a structured approach to data-driven learning. The table below outlines how each phase of CRISP-DM is adapted for cybersecurity workforce development.

Step	Description
Business Understanding	Define cybersecurity objectives and risks
Data Understanding	Collect and analyze security-related data
Data Preparation	Clean and structure data for modeling
Modeling	Develop predictive security models
Evaluation	Assess model performance and accuracy
Deployment	Implement models into security workflows
Cybersecurity Training Integration	Apply insights for workforce upskilling

Figure 3: CRISP-DM Pipeline for Cybersecurity Training

CRISP-DM is a widely recognized methodology that outlines a structured approach to data mining and analysis. It consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. By applying CRISP-DM to cybersecurity threat modeling, professionals can systematically analyze attack patterns, simulate potential threats, and develop effective defense strategies. This structured approach ensures comprehensive coverage of various aspects of threat analysis and mitigation [16].

Jupyter Notebooks for Interactive Training

Jupyter Notebooks offer an interactive computing environment that combines live code, equations, visualizations, and narrative text. In the context of cybersecurity training, Jupyter Notebooks enable professionals to explore security datasets, develop scripts for threat detection, and visualize attack patterns in a collaborative setting. This interactive approach enhances the learning experience by allowing hands-on experimentation and immediate feedback [17].

Terraform-Based Immersive Labs for Cloud Security

Terraform is an Infrastructure as Code (IaC) tool that allows for the automation of cloud infrastructure deployment. By

utilizing Terraform to create immersive lab environments, cybersecurity professionals can simulate real-world cloud attack scenarios and practice incident response strategies. This hands-on experience is crucial for understanding the complexities of cloud security and developing effective defense mechanisms.

Aligning Training with CyberSeek Heatmap Metrics

To ensure that the training framework addresses the most critical skill gaps in the cybersecurity workforce, it is essential to align the curriculum with insights from the CyberSeek Heatmap. The CyberSeek Heatmap provides detailed information on supply and demand for cybersecurity skills across various regions and roles. By leveraging this data, training programs can be tailored to focus on high-demand skills, thereby enhancing the employability of professionals and strengthening organizational security postures.

Implementing this integrated training framework will bridge the gap between theoretical knowledge and practical application, equipping cybersecurity professionals with the necessary skills to effectively combat evolving cyber threats.

5. Conclusion

Cybersecurity workforce upskilling is critical in mitigating evolving threats, and automation-driven approaches provide an effective means of enhancing training methodologies. By leveraging CRISP-DM frameworks with Jupyter Notebooks, professionals can streamline threat modeling, automate security analysis, and gain hands-on experience with real-world attack scenarios. Additionally, Immersive Lab Terraform modules allow learners to engage in practical cloud security exercises, reinforcing key competencies required for modern cybersecurity roles.

The integration of CyberSeek Heatmap metrics ensures that training aligns with industry demands, addressing workforce skill gaps and enhancing job market readiness. As cyber threats grow increasingly complex, adaptive training solutions that incorporate automation, AI-driven analytics, and cloud-based simulations will be essential in equipping professionals with the necessary skills to safeguard digital infrastructures. Continued innovation in cybersecurity education will play a pivotal role in developing a resilient and well-prepared workforce capable of responding to emerging security challenges.

References

[1] "CRISP-DM – Machine Learning Bookcamp, " *Machine Learning Bookcamp*, 2025. Available: https: //mlbookcamp. com/article/crisp-dm?

[2] J. Clinton, T. Khabaza, T. Reinartz, and R. Wirth, "The CRISP-DM Process Model, " 1999. Available: https: //keithmccormick. com/wp-content/uploads/CRISP-DM%20No%20Brand. pdf

[3] R. Holloway, A. Filos-Ratsikas, and A. Hollender, "Two's Company, Three's a Crowd: Consensus-Halving for a Constant Number of Agents * Argyrios Deligkas. " Available: https: //arxiv. org/pdf/2007.15125. pdf.

[4] P3tra-WP, "GitHub-P3tra-WP/Jupyter-Threat, " *GitHub*, 2020. Available: https: //github. com/P3tra-WP/Jupyter-Threat.

[5] "Cloud Security Training-Immersive, " *Immersivelabs. com*. Available: https: //www.immersivelabs. com/products/cloud-security?

[6] E. Aldridge, "Cybersecurity Skills Gap: Strategies for a Secure Future, " *Educate 360 Professional Training Partners*, Jul.31, 2024. Available: https: //educate360. com/blog/cybersecurity-skills-gap/?

[7] Quzara LLC, "Cybersecurity Analysis with Jupyter, " *Quzara. com*, May 23, 2024. Available: https: //quzara. com/blog/cybersecurity-analysis-with-jupyter?

[8] A. Lee, K. King, D. Gračanin, and M. Azab, "Experiential Learning Through Immersive XR: Cybersecurity Education for Critical Infrastructures, " *Lecture notes in computer science*, pp.56–69, Jan.2024, doi: https: //doi. org/10.1007/978-3-031-61382-1_4

[9] Y. Zhou and Y. Lu, "Multiple Instance Learning with Task-Specific Multi-Level Features for Weakly Annotated Histopathological Image Classification, " *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.1366–1370, May 2022, doi: https: //doi. org/10.1109/icassp43922.2022.9747121. Available: https: //ieeexplore. ieee. org/document/9747121.

[10] "Cybersecurity Supply And Demand Heat Map, " *Cyberseek. org*, 2023. Available: https: //www.cyberseek. org/heatmap. html?.

[11] HROlive, "GitHub-HROlive/Applications-of-AI-for-Anomaly-Detection: Nvidia DLI workshop on AI-based anomaly detection techniques using GPU-accelerated XGBoost, deep learning-based autoencoders, and generative adversarial networks (GANs) and then implement and compare supervised and unsupervised learning techniques., " *GitHub*, 2023. Available: https: //github. com/HROlive/Applications-of-AI-for-Anomaly-Detection?.

[12] "Morpheus Cybersecurity Solutions, " *NVIDIA Developer*, 2025. Available: https: //developer. nvidia. com/morpheus-cybersecurity?

[13] Mst. Alema Khatun and M. Abu Yousuf, "Human Activity Recognition Using Smartphone Sensor Based on Selective Classifiers," *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, pp.1–6, Dec.2020, doi: https: //doi. org/10.1109/sti50764.2020.9350486. Available: https: //ieeexplore. ieee. org/document/9350486.

[14] N. Ramsey, "Beyond Relooper: recursive translation of unstructured control flow to structured control flow (functional pearl)," *Proceedings of the ACM on Programming Languages*, vol.6, no. ICFP, pp.1–22, Aug.2022, doi: https: //doi. org/10.1145/3547621

[15] W. Newhouse, S. Keith, B. Scribner, and G. Witte, "National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework," *National Initiative for Cybersecurity Education (NICE) Cybersecurity Workforce Framework*, Aug.2017, doi: https: //doi. org/10.6028/nist. sp.800-181. Available: https: //nvlpubs. nist. gov/nistpubs/SpecialPublications/NIST. SP.800-181. pdf

[16] N. Hotz, "What is CRISP DM?-Data Science PM," *Data Science PM*, Sep.10, 2018. Available: https: //www.datascience-pm. com/crisp-dm-2/?.
[17] "Jupyter Notebooks — Threat Hunter Playbook," *Threathunter playbook. com*, 2022. Available: https: //threathunterplaybook. com/tutorials/jupyter/introduction. html?