Problem Statement: Software Datathon

1. Software Vulnerability in real-world Dataset

- **Specific Definition**:
  This dataset focuses on vulnerabilities like buffer overflow, SQL injection, and cross-site scripting (XSS). It provides detailed data on each vulnerability, such as severity, impact, and mitigation strategies, enabling the development of automated vulnerability detection tools for software applications.
- **Use Case**:
  Organize a **cybersecurity workshop** where participants analyze real-world software vulnerabilities from this dataset. They will create or refine tools for detecting and mitigating vulnerabilities, highlighting the importance of secure coding practices and ethical hacking.
- **Dataset Link**:
  https://ieee-dataport.org/documents/some-software-vulnerability-real-world-data-sets

2. Machine Learning Datasets for Software Applications

- **Specific Definition**:
  This dataset provides data suitable for training machine learning models to detect bugs, assess code quality, and optimize performance. It contains historical data on software defects and usage patterns, ideal for building models that enhance software development workflows.
- **Use Case**:
  Host a **machine learning workshop** where participants create models to predict software bugs or optimize code performance using this dataset. The workshop will showcase the power of machine learning in improving software quality and efficiency. ● **Dataset Link**:
  https://ieee-dataport.org/documents/medical-imaging-datasets-multimodal-disease-detecti on-and-diagnosis-research

3. Data Preprocessing Tools (Reinforced BCI)

- **Definition:**

  This dataset includes **reinforcement learning models** and **BCI** signals, designed for use in **brain-computer interfaces**. It provides data to train models that can **predict** or **control actions** using **neural signals**, useful for BCI applications like assistive technology or **neurofeedback systems**.

- **Use Case:**

  Host a workshop on **Brain-Computer Interfaces** where participants use the dataset to train reinforcement learning models for controlling devices through neural signals. This event could explore how BCI technology can be applied to real-world applications such as **prosthetics**, **neurofeedback**, or **assistive technology**.

- **Dataset Link**:
  https://ieee-dataport.org/documents/reinforced-bci

4. Simulation Software Dataset (Road Traffic Data)

● **Specific Definition**:
This dataset contains data generated from simulations in engineering, physics, and system optimization. It supports testing and validating models in various scenarios like disaster management, system performance under stress, and environmental studies. ● **Use Case**:
Host a **simulation modeling seminar** where participants use the dataset to create and test simulations for different engineering or environmental scenarios. This will help attendees understand the application of simulation software in optimizing system designs or modeling real-world situations.
● **Dataset Link**:
https://ieee-dataport.org/documents/road-traffic-data

5. Software Development Metrics Dataset (Prioritizing Tasks)

● **Specific Definition**:
This dataset includes performance and quality metrics of software projects, such as lines of code, code complexity, defect rates, and developer productivity. It is essential for analyzing software development processes and identifying areas for improvement. ● **Use Case**:
Host a **data analysis workshop** where participants analyze development metrics to assess the success of software projects. They will use the dataset to optimize software development practices and enhance the effectiveness of their development teams. ● **Dataset Link**:
https://ieee-dataport.org/documents/prioritizing-tasks-software-development-systematic-li terature-review

DataThon: Online Submission Guidelines

(Software Vulnerability in Real-World Dataset)

What to Submit?

- Your submission must align with the theme of the DataThon (**Human Action Recognition and Its Applications**).
- The project should address security risks and leverage the provided dataset for analysis.

- **Each team must submit:**
  - A **demo video** that showcases the flow of your idea, key implementation phases, and project functionality.
  - A **PowerPoint presentation (PPT)** covering:
    - **Project Overview** – Define the problem and your proposed solution. ▪ **Approach & Methodology** – Describe frameworks, tools, and techniques used.
    - **Dataset Used** – Specify the dataset provided and how you processed it.
    - **Key Findings & Results** – Highlight insights derived from your analysis.
    - **Conclusion & Future Scope** – Discuss potential improvements and scalability.
    - **Team Details** – Names & email addresses of all members.

PPT Guidelines: Maximum 6 slides, including team details and conclusion.

How to Submit?

- A Google Drive link will be shared to upload the files.
- Ensure your submission is uploaded **before the deadline (9th February 2025)** to be eligible.
- Follow this naming format for your files:
  - **Video:** *TeamName_ProjectName.mp4*
  - **PPT:** *TeamName_ProjectName.pptx*

Selection & Next Steps…

• A panel of judges will evaluate all submissions.
• Evaluation of the submitted documents will be done by **12th of February 2025**. • Shortlisted teams will receive email notifications.
• Top-performing teams will advance to the next stage of the competition, where they will compete against other teams live on the day of the event at SCET. • Final winners will be selected based on their live performance.

Need Help?

For any questions, reach out to:

Aakash Jaiswal - 7041848454

Himanshi Borad – 9924852451

Analyze, Secure, Innovate – Compete for the top spot!