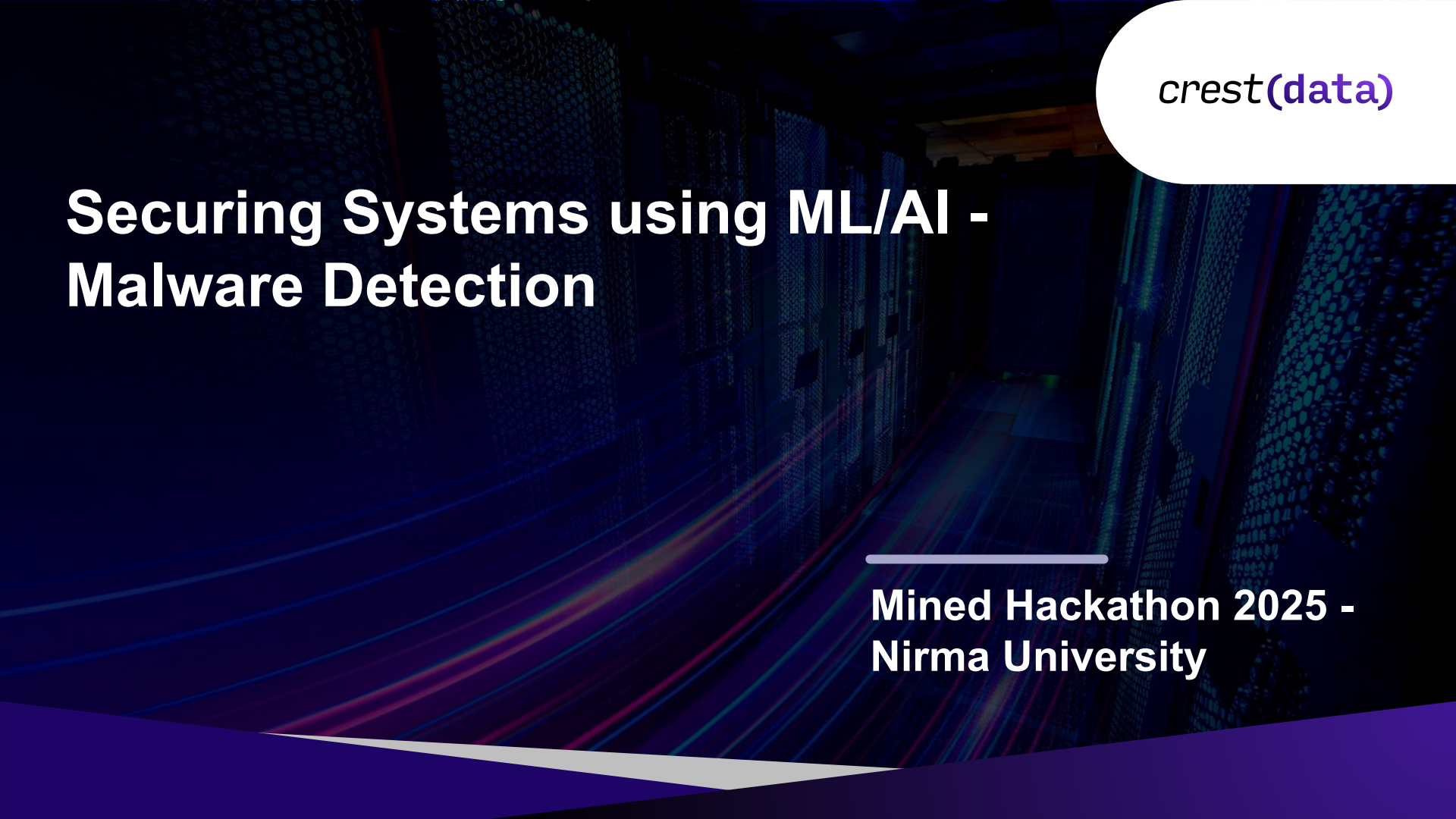

crest(data)

Securing Systems using ML/AI - Malware Detection

Mined Hackathon 2025 - Nirma University

Agenda

- Objective
- Challenges
- Dataset explanation
- Deliverables

Crest Data Systems Confidential. Do NOT Distribute.

Objective

- Participants are tasked with building a machine learning model to classify malware into predefined categories using the provided dataset.

- The goal is to develop efficient system that can accurately identify the type of malware based on the given data.

- The solution will help enhance the detection and prevention mechanisms for real-world cybersecurity systems.

Crest Data Systems Confidential. Do NOT Distribute.

Challenges

Participants must design, train, and evaluate a machine learning model to classify malware into one of the following categories:

- Benign = 0
- RedLineStealer = 1
- Downloader = 2
- RAT = 3
- BankingTrojan = 4
- SnakeKeyLogger = 5
- Spyware = 6

The implementation must be robust and scalable, with a focus on generalizing to unseen malware samples.

Crest Data Systems Confidential. Do NOT Distribute.

Dataset Explanation

The overall features are distributed in three sections:

- Portable executable: It contains 52 fields of PE headers, 9 field values of 10 PE section,
- DLL imported: contains the DLLs imported by each malware family.
- API functions: contains the API functions called by these malware

Dataset: https://drive.google.com/drive/folders/17BKEb8ujyf1lpX2hHCcXrl2Zc7mb2Vmp?usp=drive_link

Crest Data Systems Confidential. Do NOT Distribute.

Dataset Explanation

To understand the different part of dataset:

Portable executable:

- https://stixproject.github.io/data-model/1.2/WinExecutableFileObj/DOSHeaderType/
- https://learn.microsoft.com/en-us/windows/win32/debug/pe-format
- https://0xrick.github.io/win-internals/pe5/#sections-and-section-headers

DLL imported:

- It represents 629 type of DLL files are used for the respective executable.

API functions:

- It represents 21918 type of API function call done from the respective executable.

Crest Data Systems Confidential. Do NOT Distribute.

Deliverables

- The machine learning model and its implementation code.
- A detailed report explaining:
  - Model architecture and approach.
  - Preprocessing and feature engineering techniques used.
  - Challenges faced and how they were overcome.
- A presentation of your solution, including key findings and potential areas for improvement.
- Test.csv along with predictions
- Trained model file

Crest Data Systems Confidential. Do NOT Distribute.

THANK YOU!

crest(data)

info@crestdata.ai
https://www.crestdata.ai