Advanced AI-Powered Adaptive Defense Against Human-Operated Ransomware

Supervised by

Dr. Muhammad Yahya Mahmoud (Assistant Professor, COE, KFUPM)
Dr. Ashraf Mahmoud (Associate Professor, COE, KFUPM)
Dr. Farid Binbeshr (Post Doctoral Fellow, Intelligent Secure Systems, KFUPM)

Presented by
Ransomware-1 Team-1

Md Siddiqur Rahman Tanveer (g202417180)
F.M. Jahiduzzaman (g202417000)

Department of Computer Engineering
King Fahd University of Petroleum & Minerals (KFUPM)

Presentation Outline

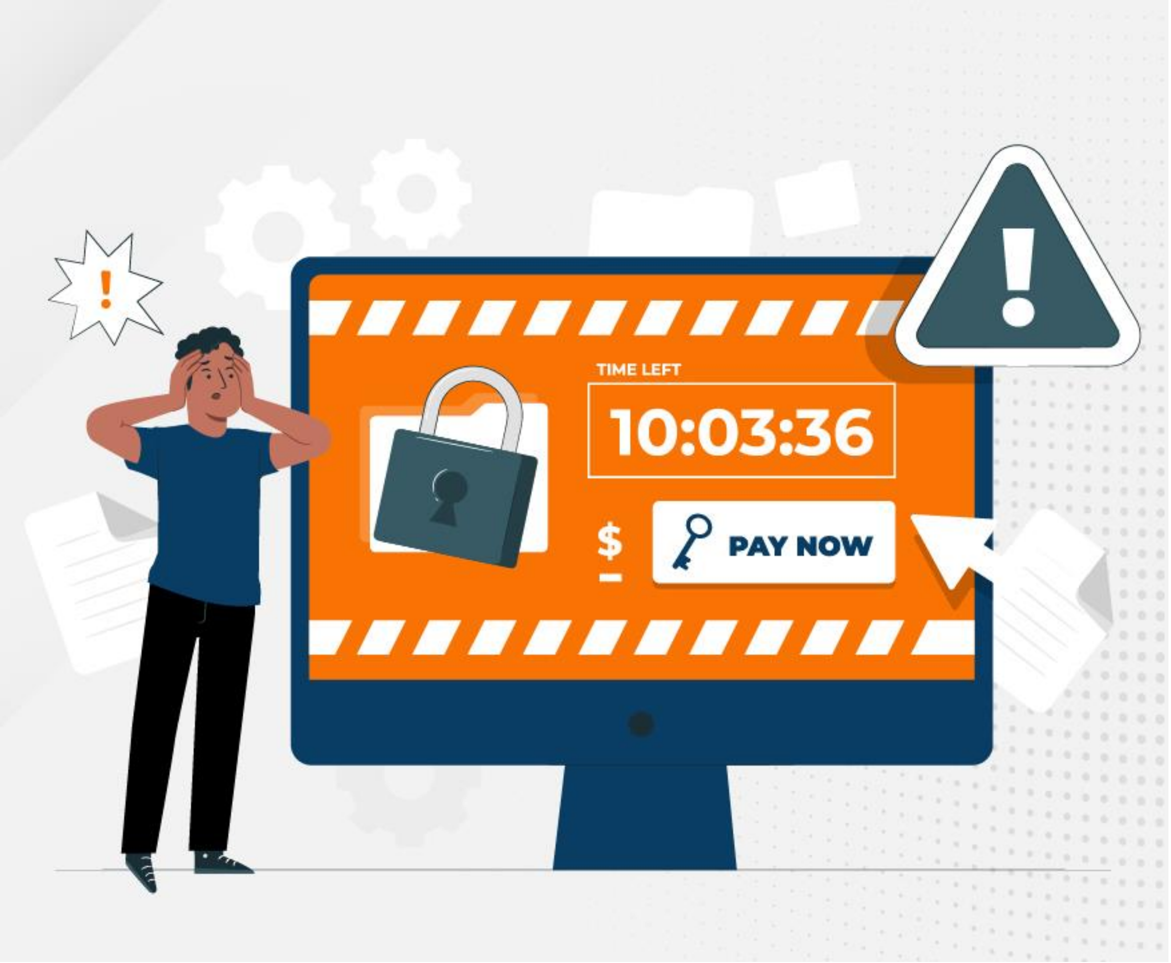
- Introduction
- Problem Statement and Motivation
- Methodology and Approach
- Implementation Details
- Results and Performance Analysis
- Conclusions and Recommendations

What is a Ransomware?

Ransomware is a type of *malware* that holds a victim's sensitive data or device hostage, threatening to keep it locked—or worse—unless the victim pays a ransom to the attacker [1]

Common Ransomware Types

- ❑ Crypto Ransomware or Encryptors
- ❑ Lockers
- ❑ Scareware
- ❑ Doxware or Leakware



[1] https://www.ibm.com/topics/ransomware

What is Human-operated ransomware?

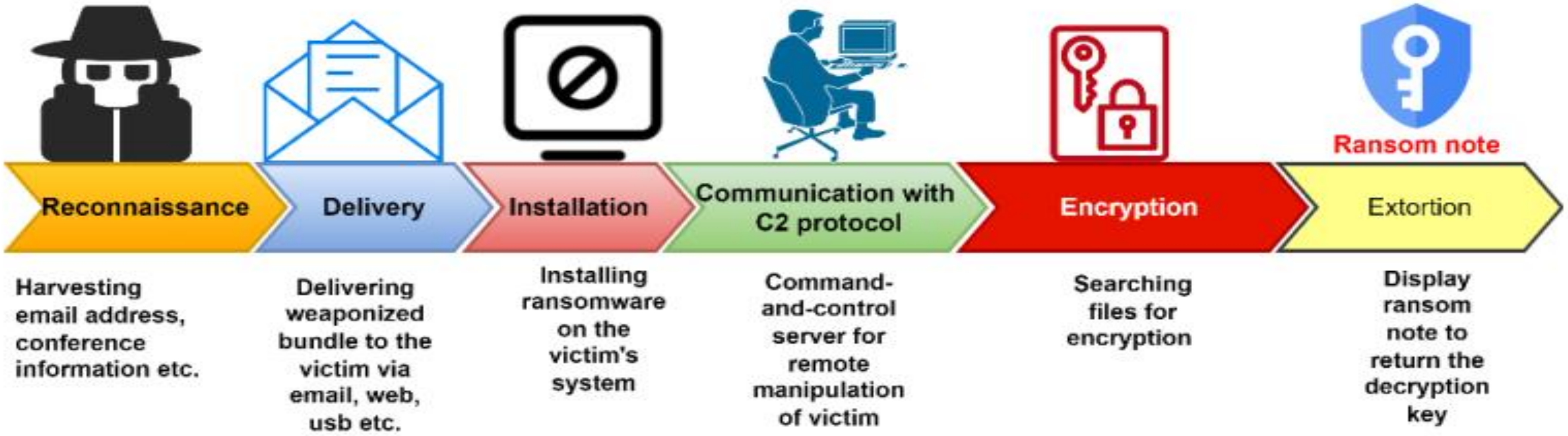
Human-operated ransomware is a planned and coordinated attack by active cybercriminals who employ multiple attack methods [2]

involves cybercriminals actively [3] like

- "**hands-on-keyboard**" operations
- focusing on entire organizations

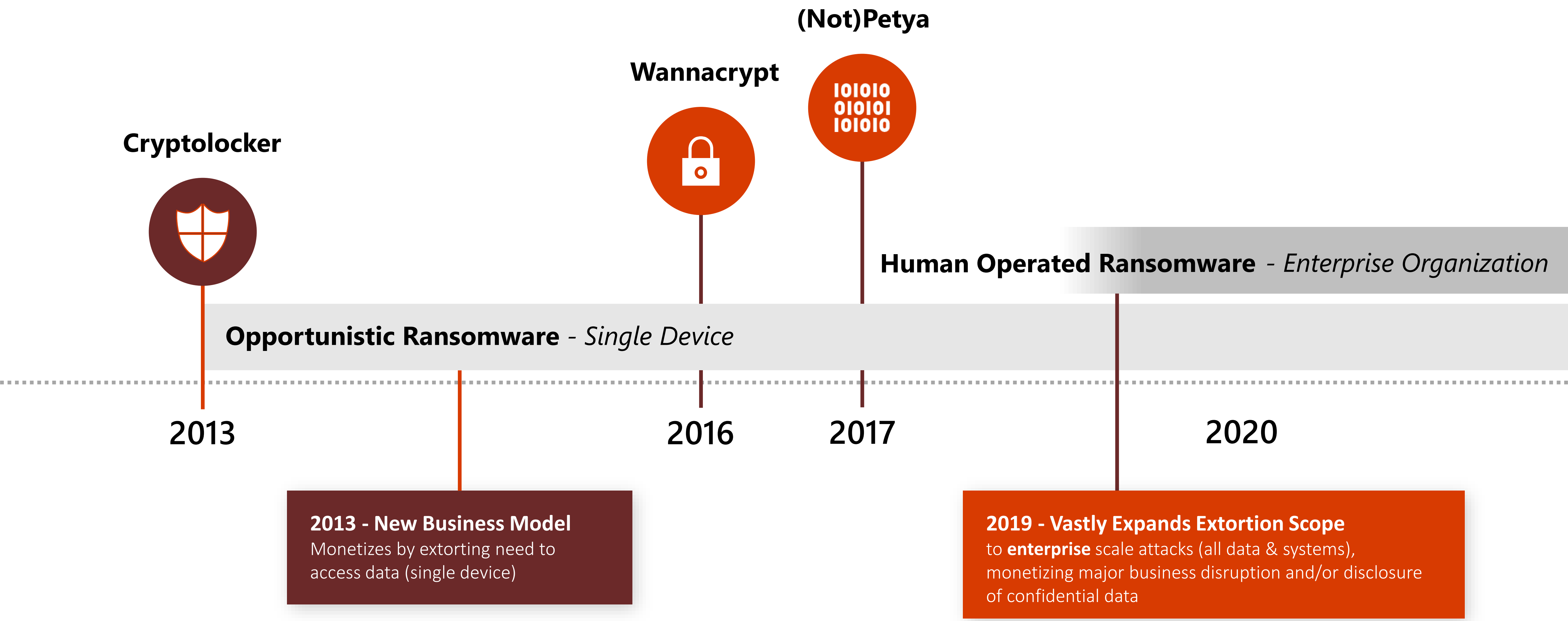
The attackers **leverage their knowledge** of

- typical system and security misconfigurations
- aim to penetrate the organization
- move laterally across the network
- exploit vulnerabilities.



[2] https://learn.microsoft.com/en-us/defender-xdr/playbook-detecting-ransomware-m365-defender
[3] Ferdous, Jannatul, Rafiqul Islam, Arash Mahboubi, and Md Zahidul Islam. "AI-based Ransomware Detection: A Comprehensive Review." *IEEE Access* (2024).

Evolution of Ransomware models [4]



[4] https://learn.microsoft.com/en-us/security/ransomware/human-operated-ransomware

Human Operated Ransomware - high impact & growing another background security risk [4]

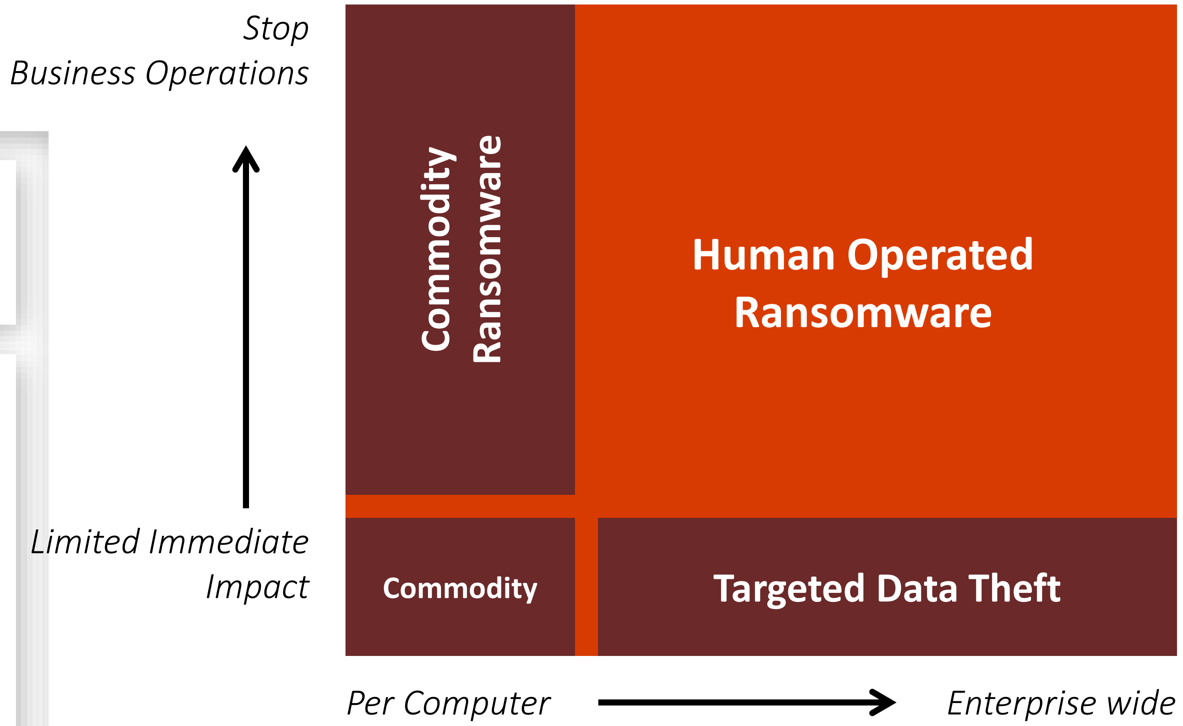
What’s different?

High Business impact
Extortion must disrupt business operations to motivate payment

Profitable for Attackers
Economic incentive to continue growing

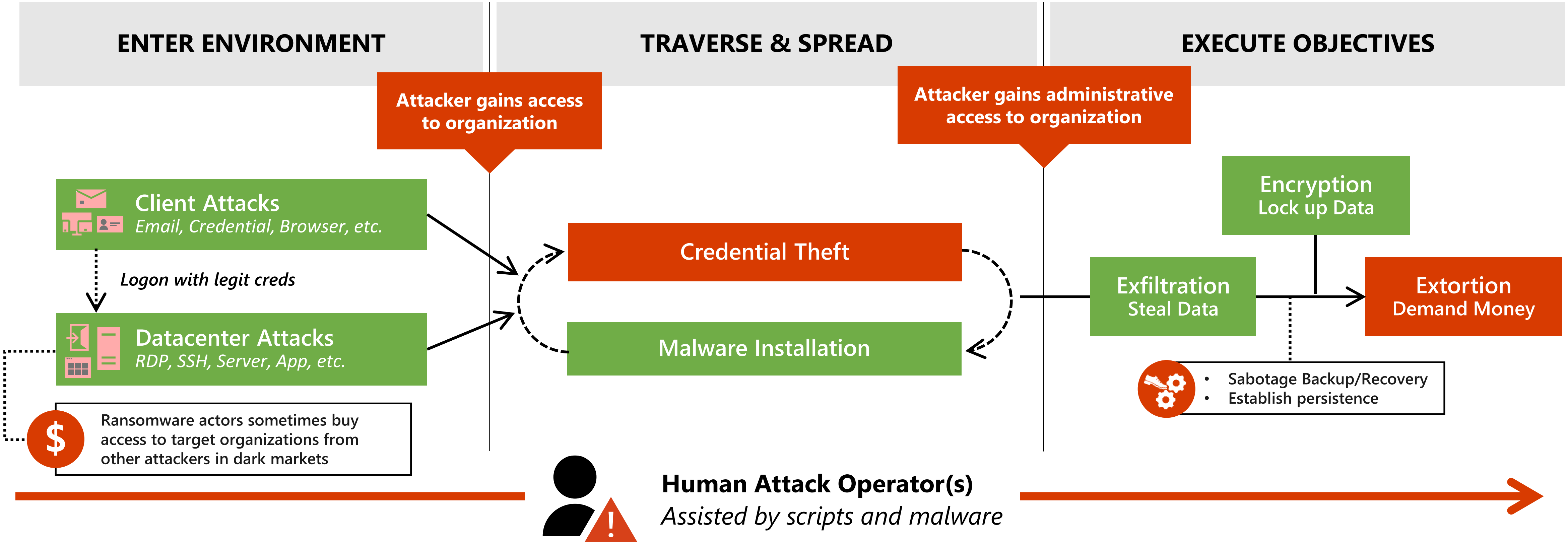
Room to Grow
Attackers can monetize security maintenance gaps at most enterprises:

- **Apply security updates** consistently to all computers
- **Securely configure all resources** using manufacturer best practices
- **Mitigate credential theft** attacks for privileged users



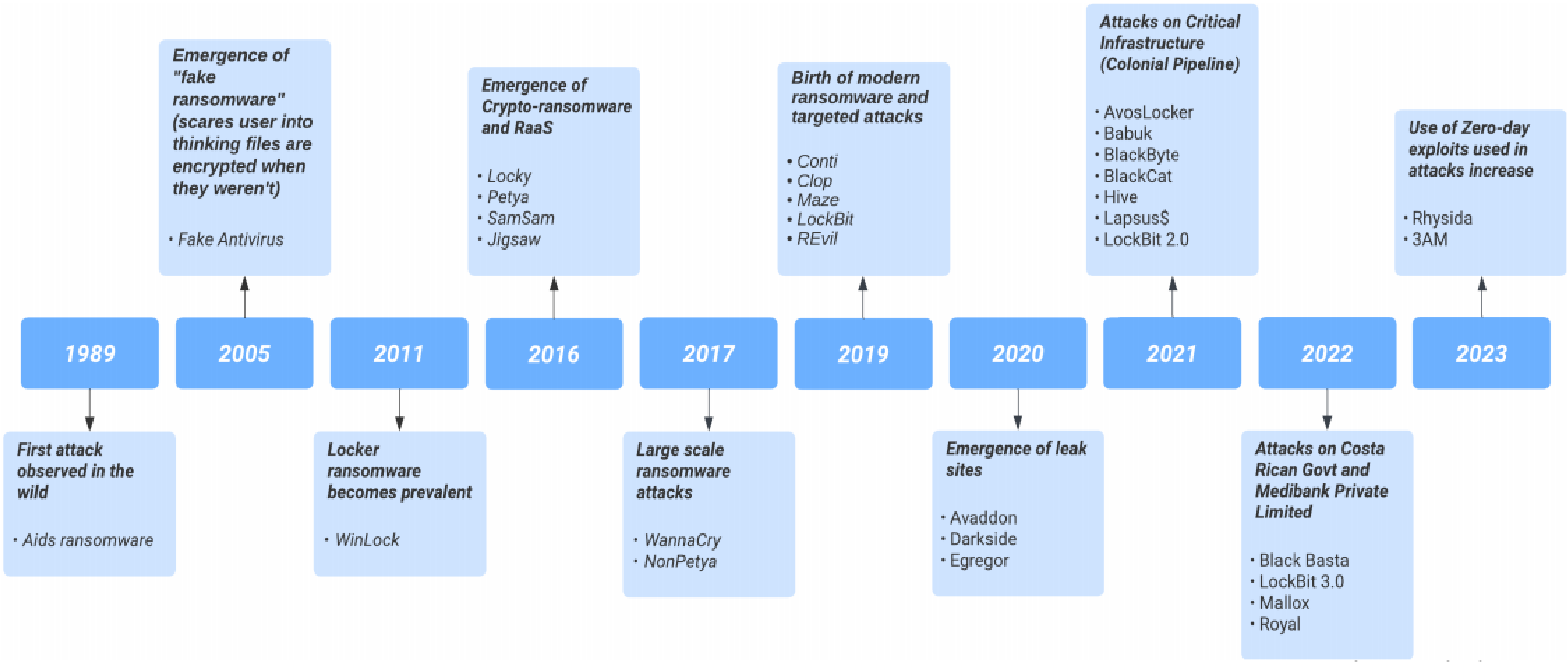
[4] https://learn.microsoft.com/en-us/security/ransomware/human-operated-ransomware

Pattern–Human Operated Ransomware[4]



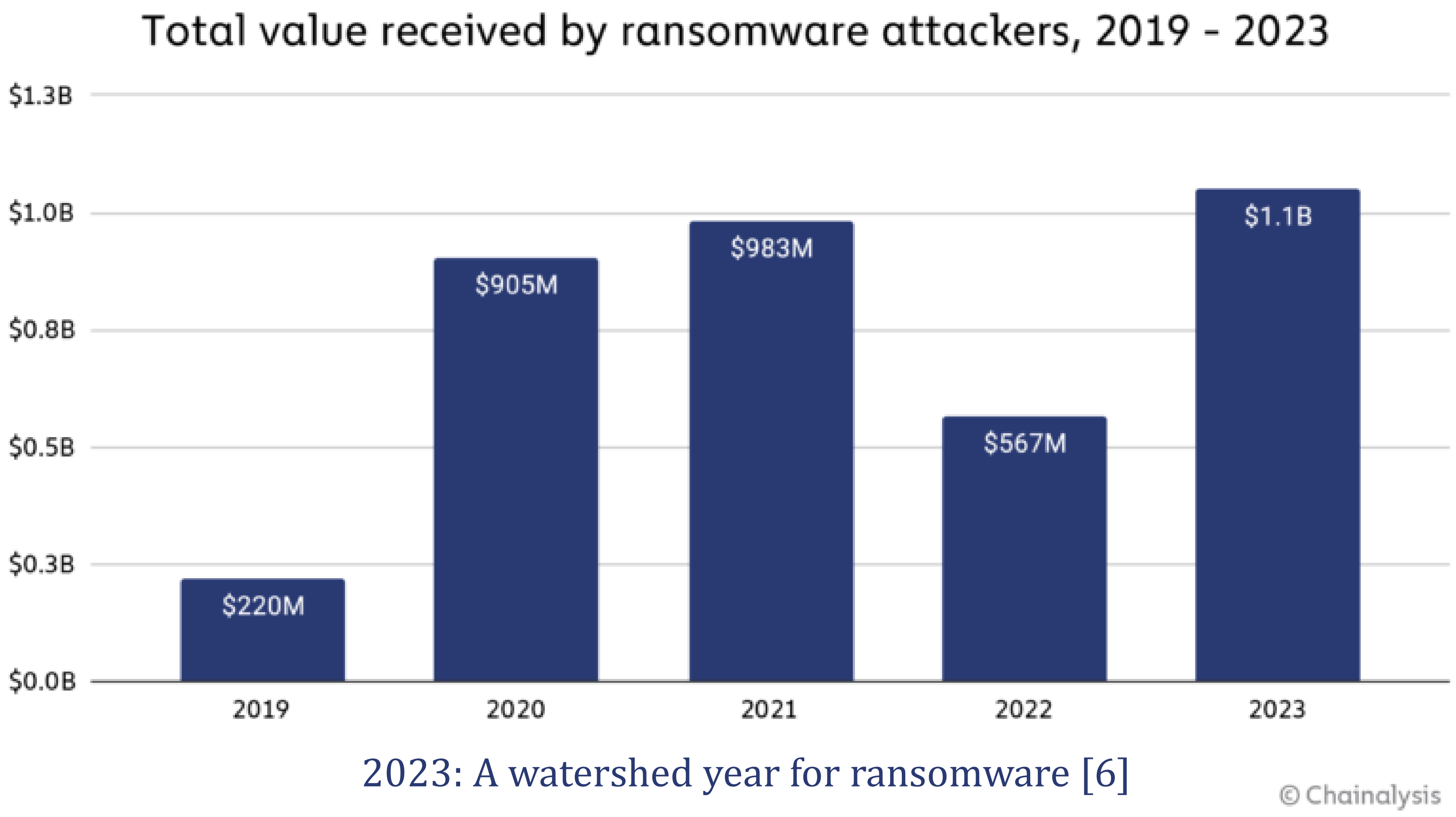
[4] https://learn.microsoft.com/en-us/security/ransomware/human-operated-ransomware

Timeline Evolution of Different Ransomwares [5]



[5] Ispahany, Jamil, MD Rafiqul Islam, Md Zahidul Islam, and M. Arif Khan. "Ransomware detection using machine learning: A review, research limitations and future directions." *IEEE Access* (2024).

Financial Loss Due to Ransomware Attacks



2023: A watershed year for ransomware [6]

[6] https://www.chainalysis.com/blog/ransomware-2024/

Most Common Human Operated Ransomware Types

[7] Alraizza, Amjad, and Abdulmohsen Algarni. "Ransomware detection using machine learning: A survey." *Big Data and Cognitive Computing* 7, no. 3 (2023): 143.

Point of Research Interest

Cybersecurity and threat actors are racing to develop advanced **AI-driven solutions** [5]. They are utilizing

Machine Learning (ML) algorithms to

- identify vulnerabilities in a target system
- exploit them to gain access
- encrypt data
- rendering them unusable until a ransom is paid.

Artificial Intelligence (AI) algorithms to

- adapt and evolve tactics based on the defenses of a target
- making it **increasingly difficult to detect** and mitigate attacks.

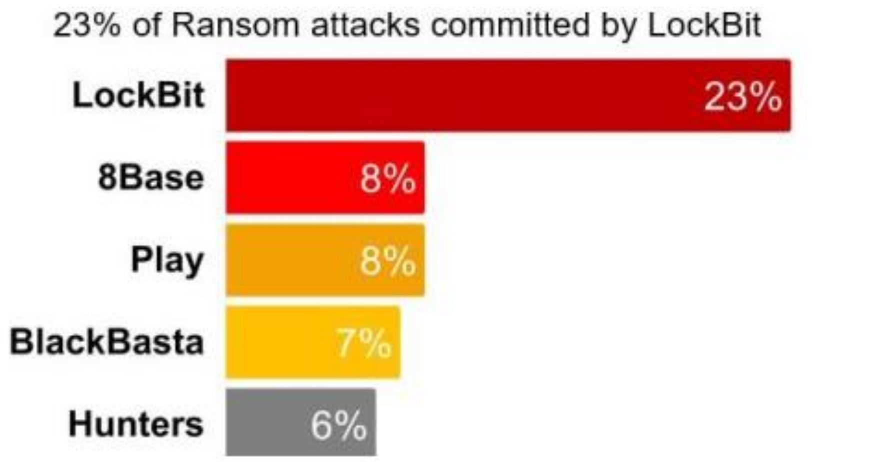


FIGURE 2. Ransomware attack distributions by groups: Q1/2024

Examples of AI-powered ransomware attacks include "LockBit 2.0", released in January 2023, which can encrypt more data and demand higher ransomware [3]

[3] Ferdous, Jannatul, Rafiqul Islam, Arash Mahboubi, and Md Zahidul Islam. "AI-based Ransomware Detection: A Comprehensive Review." *IEEE Access* (2024).

Existing Approaches to Defend Against Ransomware

Year	Authors	Proposed Solution	Model	Dataset	Samples	Features	Outcomes
2016	Kharraz et al. [8]	Dynamic analysis system called UNVEIL	Statistical	Custom Generated	148,223	30967	Total Samples: 148,223 Detected Ransomware: 13,637 Detection Rate: 96.3% False Positives: 0.0% New Detection: 9,872 (72.2%)
2021	Almousa et al [9]	API-based obfuscation techniques	k-NN SVM RF	Custom Generated	Ransomware: 58 Good: 66	206 common API Calls	K-NN : 99.18%, FP: 1 SVM: 83.60%, FP: 17 RF: 87%, FP: 12
2022	Masum et al. [10]	ML based detection	DT, RF, NB, LR, NN	Publicly available but Custom Generated	138047 70 % **Good** 30% **Ransom**	13	DT: 0.98±0.01 RF: 0.99±0.01 NB: 0.35±0.03 LR: 0.96±0.02 NN: 0.97±0.01

[8] Kharaz, A., Arshad, S., Mulliner, C., Robertson, W., & Kirda, E. (2016). {UNVEIL}: A {Large-Scale}, automated approach to detecting ransomware. In *25th USENIX security symposium (USENIX Security 16)* (pp. 757-772).

[9] Almousa, M., Basavaraju, S., & Anwar, M. (2021, December). Api-based ransomware detection using machine learning-based threat detection models. In *2021 18th International Conference on Privacy, Security and Trust (PST)* (pp. 1-7). IEEE

[10] Masum, M., Faruk, M. J. H., Shahriar, H., Qian, K., Lo, D., & Adnan, M. I. (2022, January). Ransomware classification and detection with machine learning algorithms. In 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0316-0322). IEEE.

Existing Approaches to Defend Against Ransomware

Year	Authors	Proposed Solution	Model	Dataset	Samples	Features	Outcomes
2023	Moreira et al [11]	Static analysis approach for detecting ransomware by converting PE headers into color images	CNN	Dataset 1 generated, Publicly available but Generated Dataset 2	Dataset 1: R: 10,138 B: 10,138 Dataset 2: R: 9,738 B: 9,738	1024	Dataset 1: Accuracy: 97.03% Dataset 2: Accuracy: 93.28%
2023	Zumtaugwald and Gagulic [12]	Storage Access pattern analysis	RF, XGBoost, DNN	Generated		9	XGBoost & RF: Up to 97.3% DNN: Up to 95.6%
2024	Alhaidari [13]	Mechanism utilizing memory artifacts	XGBoost, RF, LightGBM, Adaptive Boosting, Extra Tree	Generated	Ransom: 586 Benign: 579	58	XGBoost (with all features): 97.85% Random Forest (with 16 selected features): 97%

[11] Moreira, C. C., Moreira, D. C., & de Sales Jr, C. D. S. (2023). Improving ransomware detection based on portable executable header using xception convolutional neural network. Computers & Security, 130, 103265.
[12] Ransomware Detection with Machine Learning with Storage Systems, Dario Gagulic, Lynn Zumtaugwald, Siddhant Sahu, Dr. Alberto Huertas, Jan von der Assen and Dr. Roman Pletka (IBM Research Lab Zurich), University of Zurich Department of Informatics (IFI), Binzmühlestrasse 14, CH-8050 Zürich, Switzerland
[13] Aljabri, M., Alhaidari, F., Albuainain, A., Alrashidi, S., Alansari, J., Alqahtani, W., & Alshaya, J. (2024). Ransomware detection based on machine learning using memory features. Egyptian Informatics Journal, 25, 100445.

Existing Approaches to Defend Against Ransomware

Ransomware Detection Based on Data

Static Analysis

Static analysis examines binary structural properties or content to identify potential vulnerabilities. **Such as PE metadata**

Dynamic Analysis

Dynamic or behavioral analysis involves executing and observing suspicious files in a controlled environment, such as a virtual machine or emulator, to understand their behavior and actions better. Such as **Windows API Calls, Registry Operation, File operation** etc.

Dynamic analysis is necessary because some ransomware can detect virtual environments and avoid displaying malicious behaviors

Finding The Gaps

Dataset Availability

- ❑ Most Datasets are not publicly available
- ❑ Lack of original Dataset
- ❑ Most of the Datasets are generated by the authors

Identifying Ransomware behaviors

- ❑ Ransomware behaviors are very dynamic
- ❑ Nowadays, ransomware adopts evasion techniques
- ❑ Ransomware exploits zero day attacks

Detection Approach

- ❑ Static approaches let malicious file not to act that's why ransomware behaviors are difficult to identify
- ❑ Dynamic approaches let the malicious file to act in a ***quarantine environment*** so that actual behaviors are not sometimes exploited by ransomware

Research Questions

- How can AI models be effectively trained to detect and respond to human-operated ransomware attacks in real-time?
- What machine learning techniques can be employed to **dynamically** adapt to new *attack patterns* and evasion tactics used by ransomware operators?
- How can human-operated ransomware be distinguished from commodity ransomware using advanced AI techniques?

Methodology

Dataset collection

- Publicly available data sources
- Static and Dynamic behaviors dataset.

Data Preprocessing and cleaning

- Clean the dataset if there exists any null or non numeric values

Tools and Library: **Excel, Python**, Pandas, Numpy

Correlated Features Selection

- Select mostly correlated static and dynamic features

Mathematical Model: **Pearson heatmap**

Relevant Features Extraction

- Select most relevant features and drop the less relevant features

Mathematical Model: **Information value (IV), Weight of Evidence (WoE)**

Train AI Model

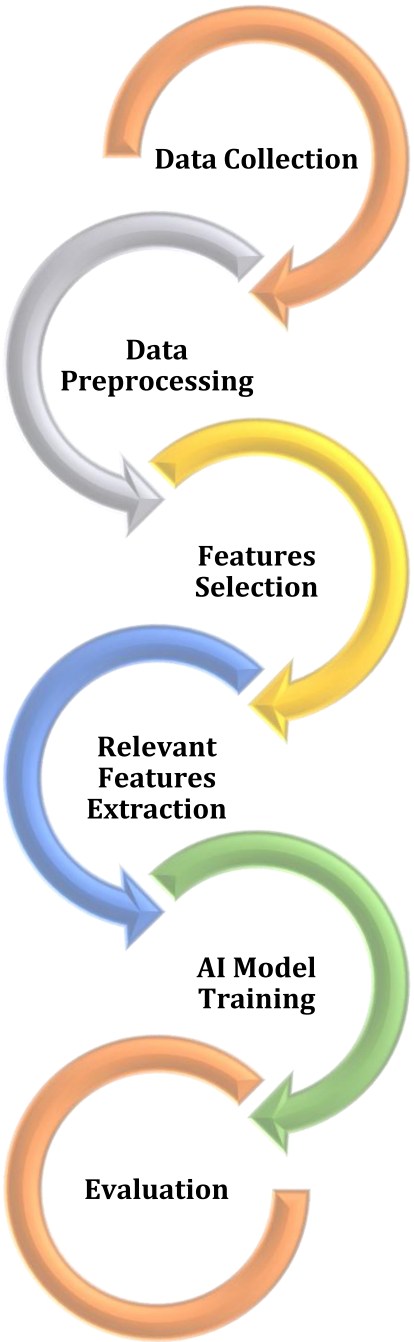
- Train several AI models using these extracted features against the dataset

Tools: **Google Colab**

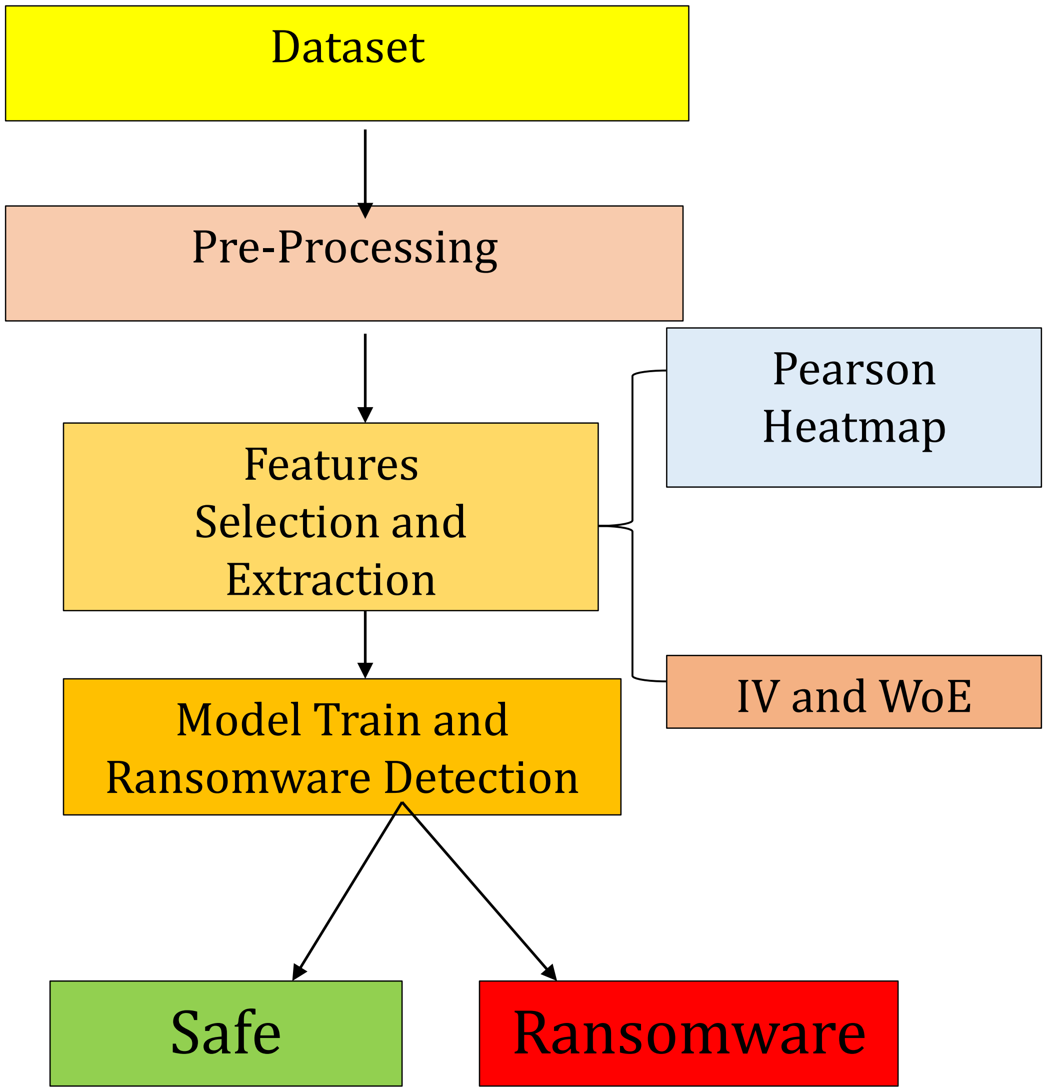
Evaluation

- Evaluate the performance by several metrics

Metrics: **Confusion matrix, Accuracy, Precision, Recall, F1 score, False Positive Rate**



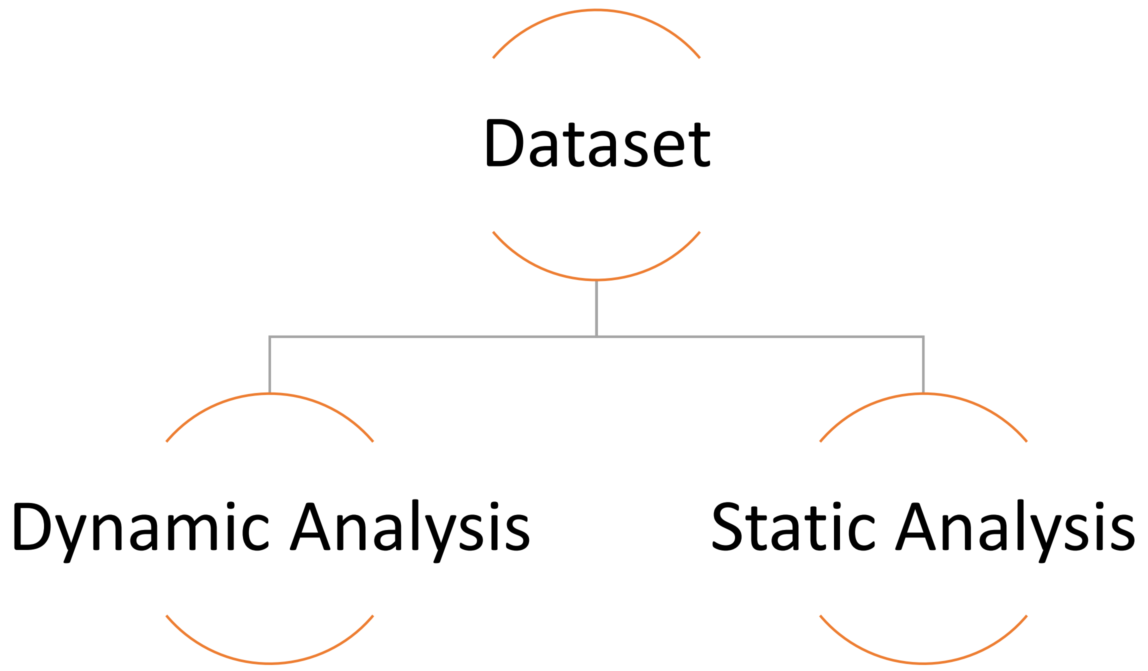
Proposed Solution



Experimental Analysis

Dataset Collection, Preprocessing and Cleaning

- Static and Dynamic behaviors dataset.

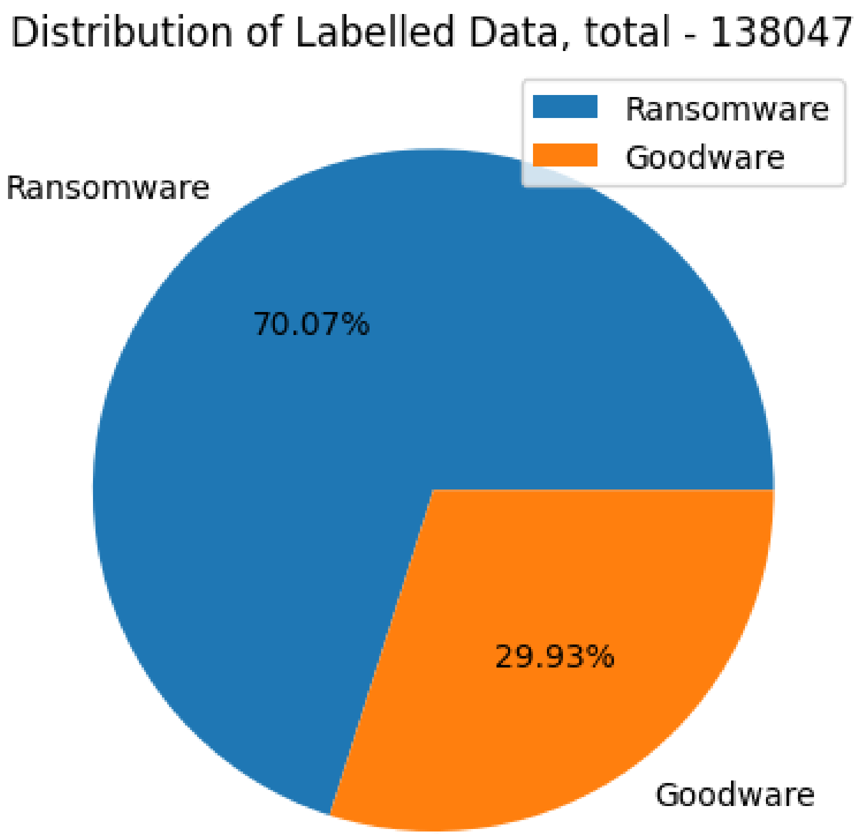


Implementation for Static Analysis

Correlated Features Selection

- Select mostly correlated static and dynamic features

Mathematical Model: **Pearson heatmap**



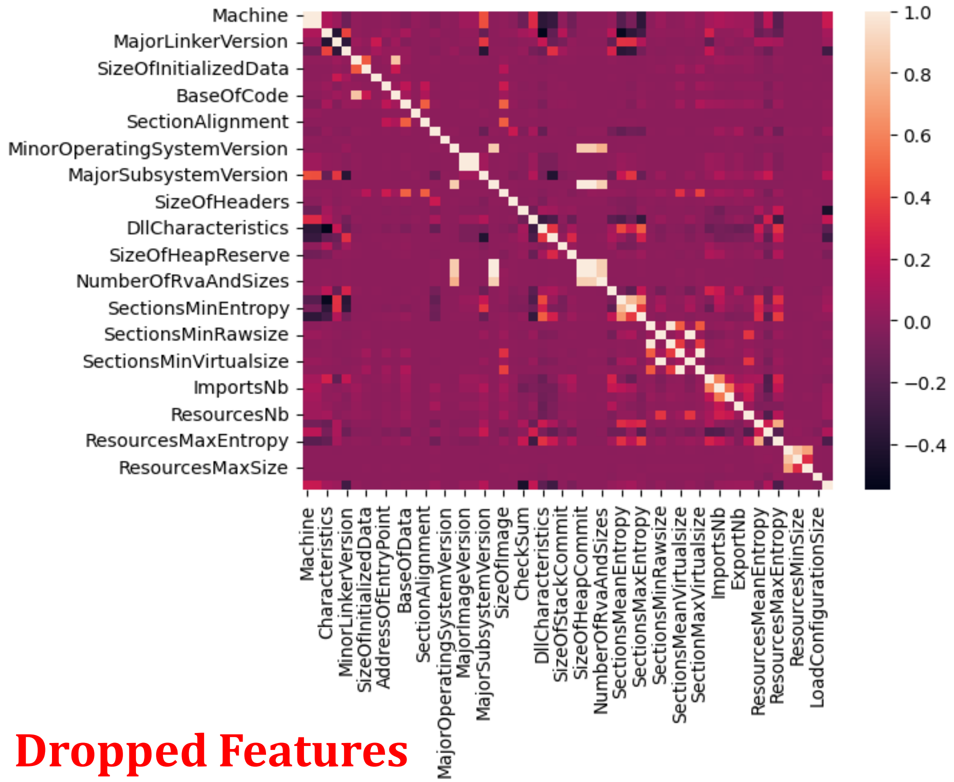
PE headers Data
Features: 54
Goodware : 41323
Ransomware : 96724

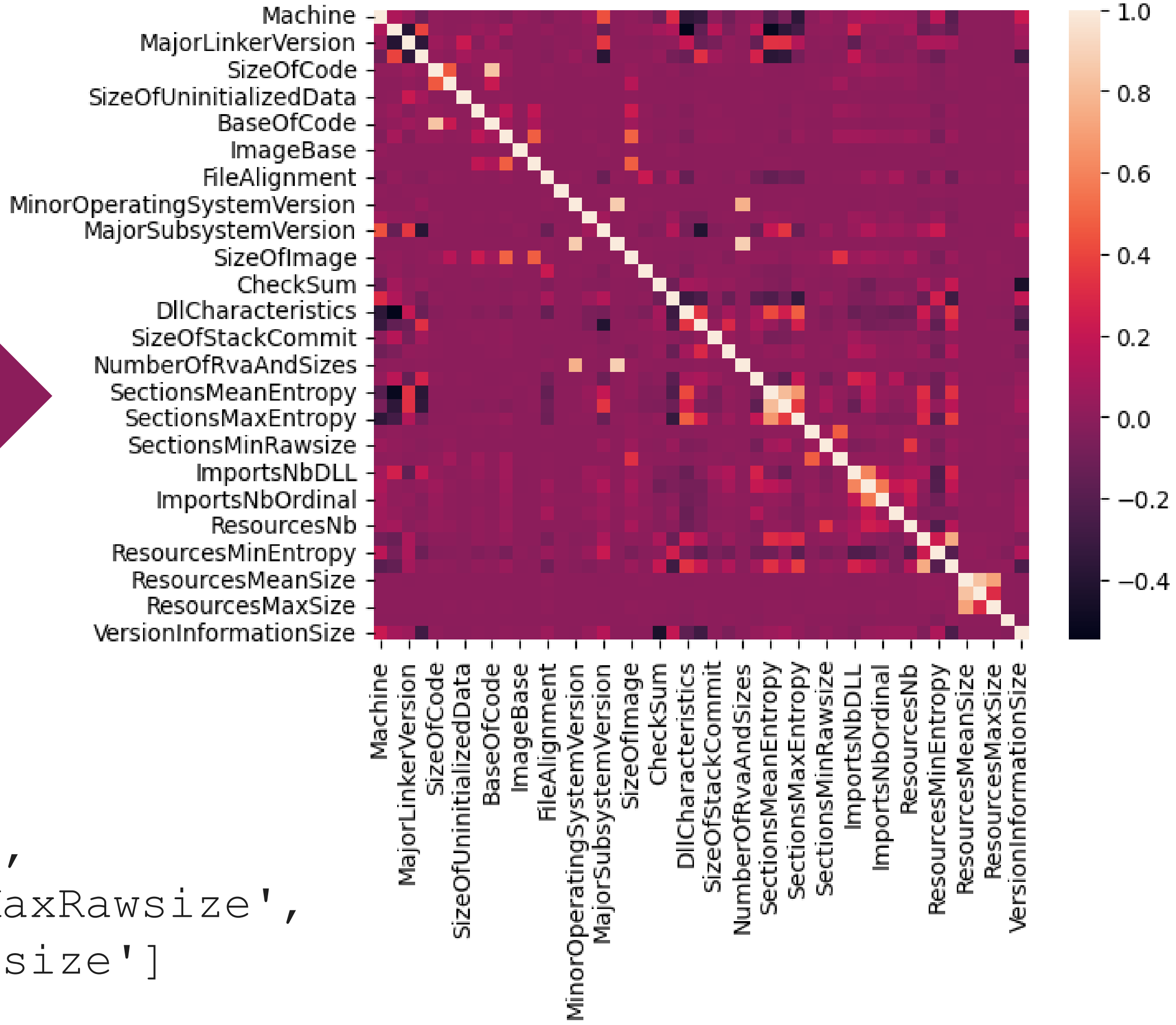
Implementation for Static Analysis

Dropping Highly Correlated Features

- Select mostly correlated static and dynamic features

Mathematical Model: **Pearson heatmap**





Dropped Features

```
['SizeOfOptionalHeader', 'MinorImageVersion',
'SizeOfHeapCommit', 'LoaderFlags', 'SectionMaxRawsize',
'SectionsMinVirtualsize', 'SectionMaxVirtualsize']
```

Implementation for Static Analysis

Relevant Features Extraction

- Select most relevant features and drop the less relevant features

Mathematical Model: **Information value (IV), Weight of Evidence (WoE)**

WOE = In(Distribution of Goods ÷ Distribution of Bads)

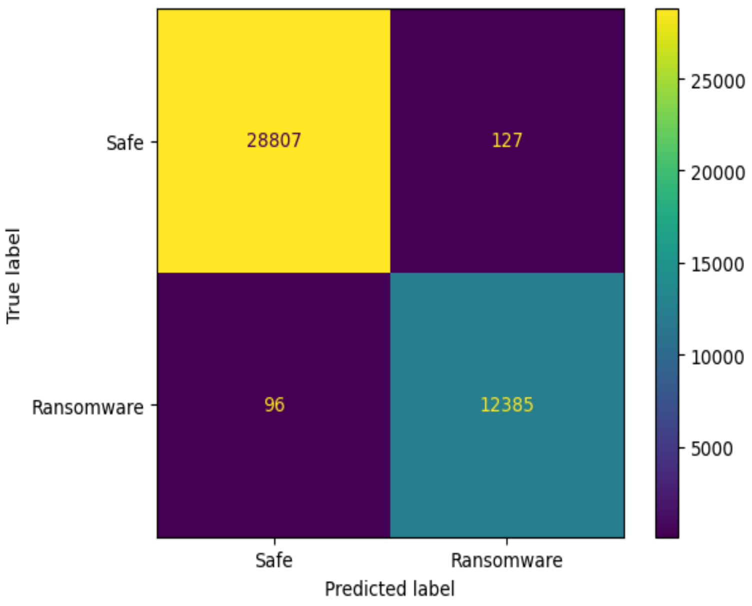
IV = ∑ (Distribution of Goods ÷ Distribution of Bads) * WOE

36 features

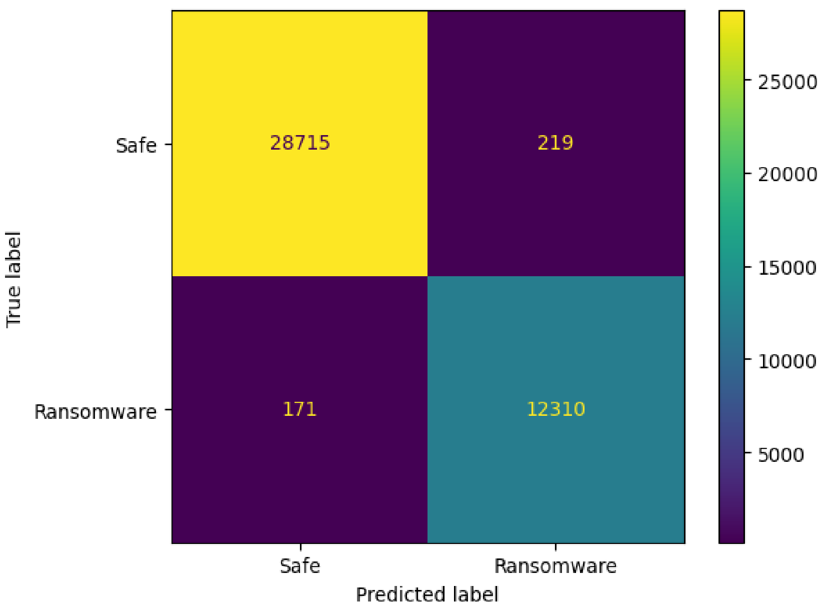
```
['SizeOfHeaders', 'FileAlignment', 'SectionAlignment', 'MajorImageVersion', 'LoadConfigurationSize',
'SizeOfUninitializedData', 'ImportsNbOrdinal', 'MinorOperatingSystemVersion', 'MinorLinkerVersion',
'MinorSubsystemVersion', 'SectionsNb', 'ResourcesMeanEntropy', 'ImportsNbDLL', 'SectionsMinEntropy',
'SectionsMeanEntropy', 'CheckSum', 'ResourcesMeanSize', 'SectionsMinRawsize', 'ResourcesMaxEntropy',
'ImportsNb', 'SectionsMeanVirtualsize', 'AddressOfEntryPoint', 'ResourcesMaxSize', 'DllCharacteristics',
'SizeOfCode', 'SectionsMeanRawsize', 'Machine', 'ExportNb', 'MajorLinkerVersion', 'SizeOfImage',
'BaseOfData', 'ResourcesMinEntropy', 'Subsystem', 'ResourcesNb', 'MajorSubsystemVersion',
'SizeOfInitializedData']
```

Train the Model for these features

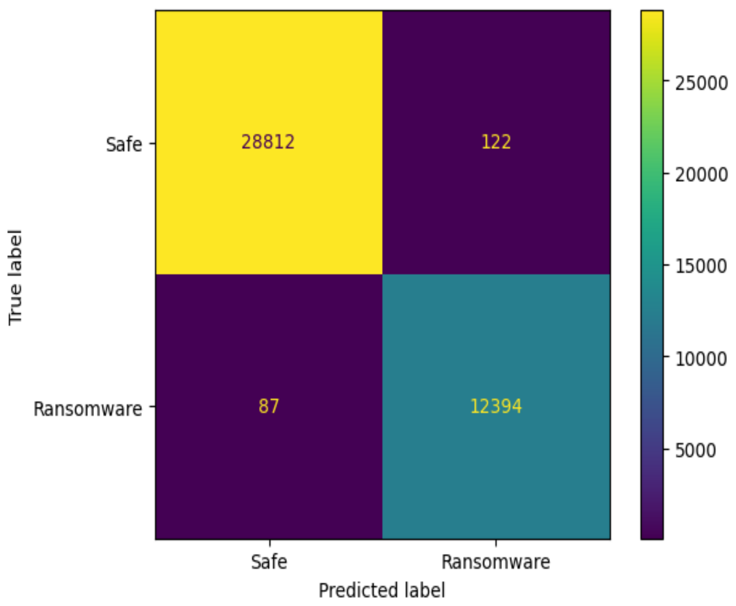
Random Forest



Decision Tree



XGBoost

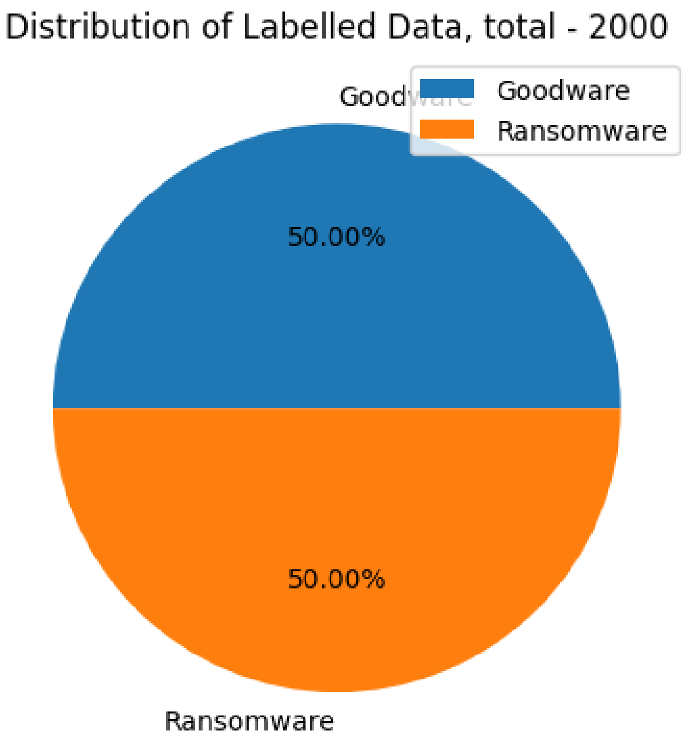


Performance Analysis For Static Analysis Dataset

Our Study	RF	DT	XGBoost		DT	RF
Accuracy	0.9946	0.9906	**0.9950**	**Masum** *et. al.* [14]	0.98±0.01	0.99±0.01
Precision	0.9898	0.9825	**0.9903**		0.98±0.00	0.99±0.00
Recall	0.9923	0.9863	**0.9930**		0.94±0.05	0.97±0.03
F1 Score	0.9911	0.9844	**0.9916**		0.94±0.05	0.97±0.03
False Positive Rate	0.0044	0.0044	**0.0042**			
AUC Score	0.9997	0.9894	**0.9998**			

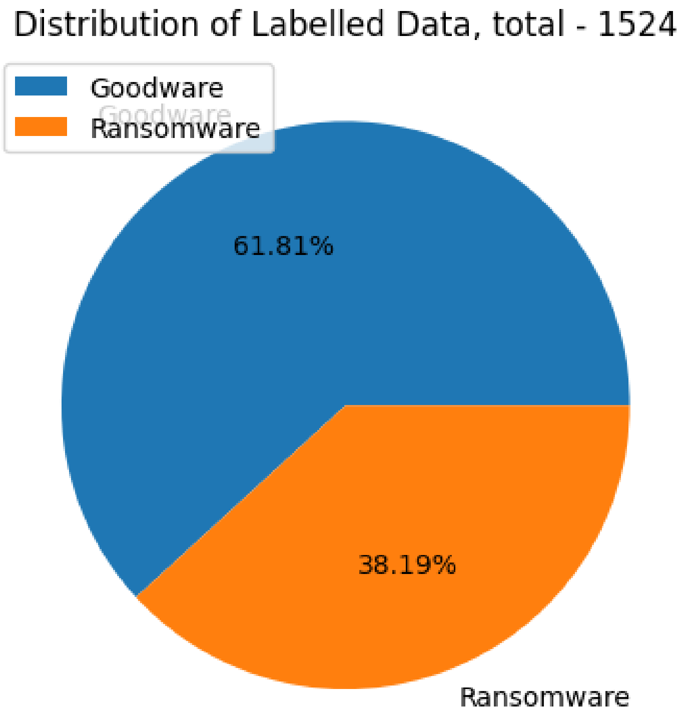
[14] Masum, Mohammad, Md Jobair Hossain Faruk, Hossain Shahriar, Kai Qian, Dan Lo, and Muhaiminul Islam Adnan. "Ransomware classification and detection with machine learning algorithms." In *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 0316-0322. IEEE, 2022.

Implementation for Dynamic Analysis



Dataset 1 Features: 50
Ransomware: 1000
Goodware: 1000

Behaviors Features Dataset



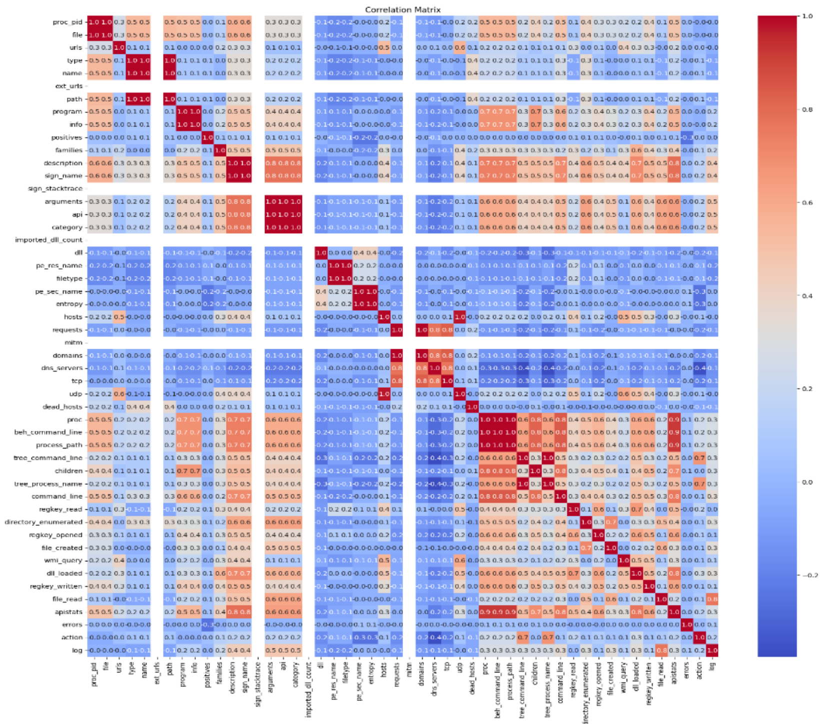
Dataset 2 Features: 30967
Ransomware: 582
Goodware: 942

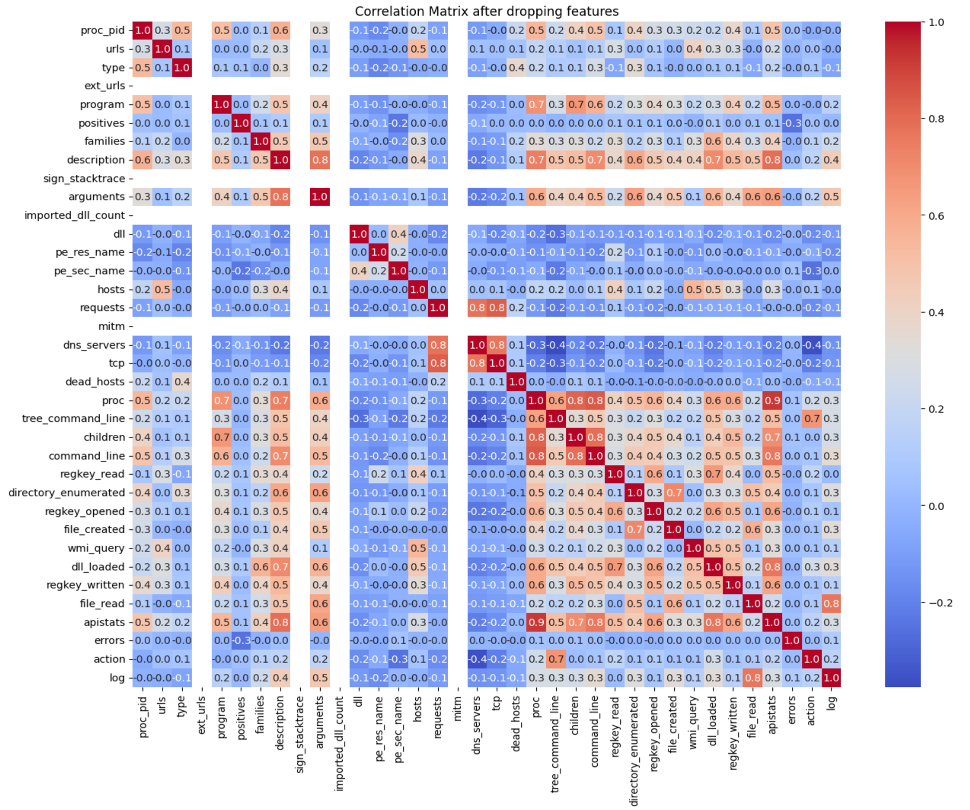
Dataset 1 Dynamic Analysis

Dropping Highly Correlated Features

- Select mostly correlated dynamic features

Mathematical Model: **Pearson correlation matrix**





Dropped Features

['file', 'name', 'path', 'info', 'sign_name', 'api', 'category', 'filetype', 'entropy', 'domains', 'udp', 'beh_command_line', 'process_path', 'tree_process_name']

Dataset 1 Dynamic Analysis

Relevant Features Extraction

- Select most relevant features and drop the less relevant features

Mathematical Model: **Information value (IV), Weight of Evidence (WoE)**

WOE = In(Distribution of Goods ÷ Distribution of Bads)

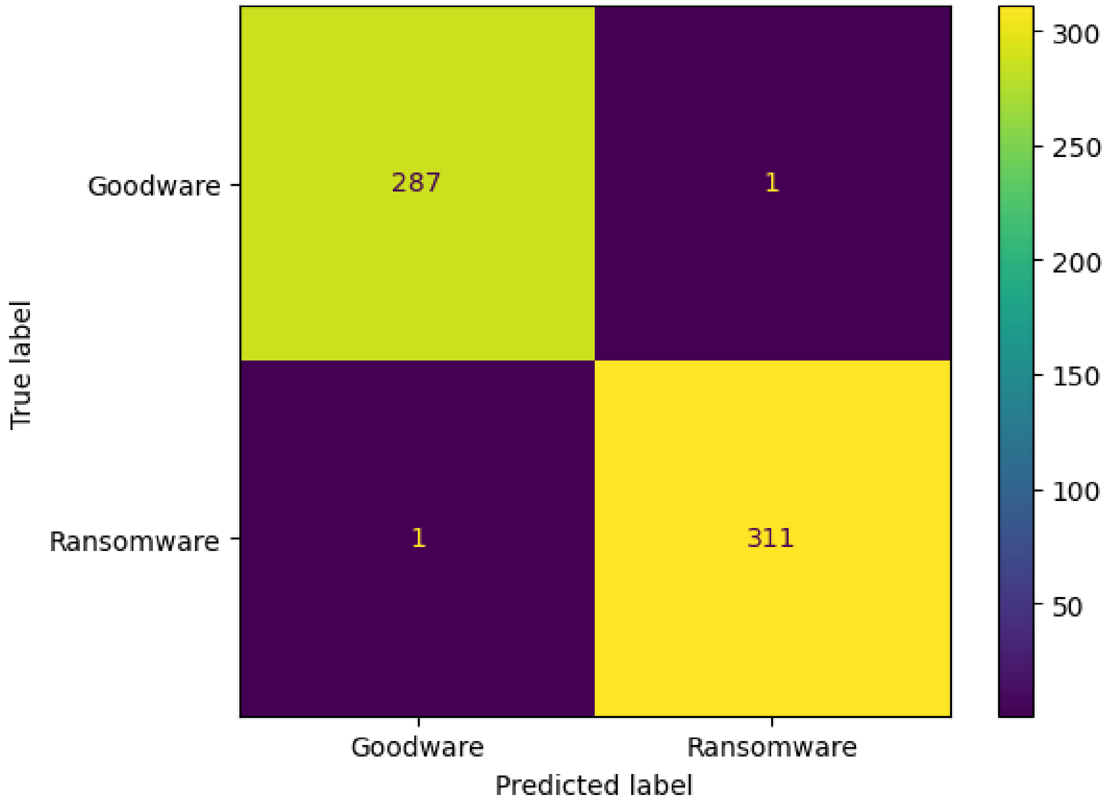
IV = ∑ (Distribution of Goods ÷ Distribution of Bads) * WOE

25 features regarding 0< IV <1.3

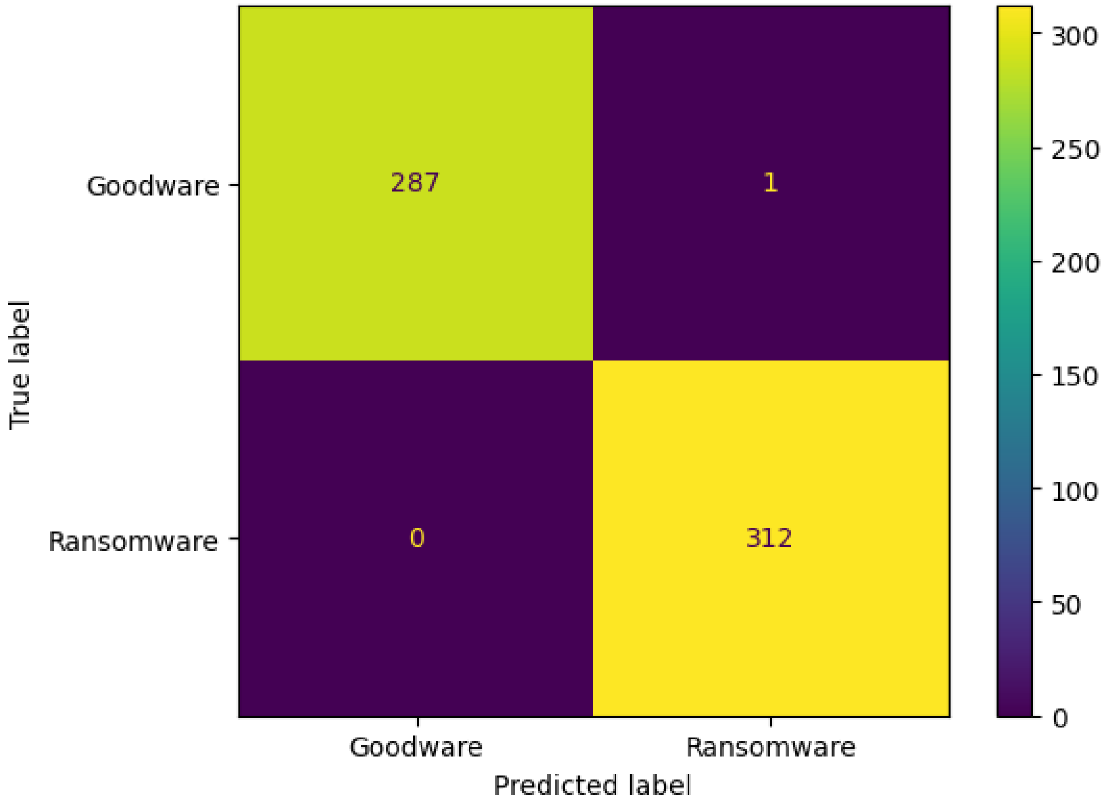
```
['positives', 'errors', 'dns_servers', 'dead_hosts', 'children', 'hosts',
'requests', 'tcp', 'action', 'regkey_opened', 'families', 'urls', 'wmi_query',
'dll_loaded', 'type', 'command_line', 'tree_command_line', 'program', 'proc',
'proc_pid', 'regkey_read', 'regkey_written', 'dll', 'apistats', 'file_read']
```

Train the Model for Dataset 1

Random Forest

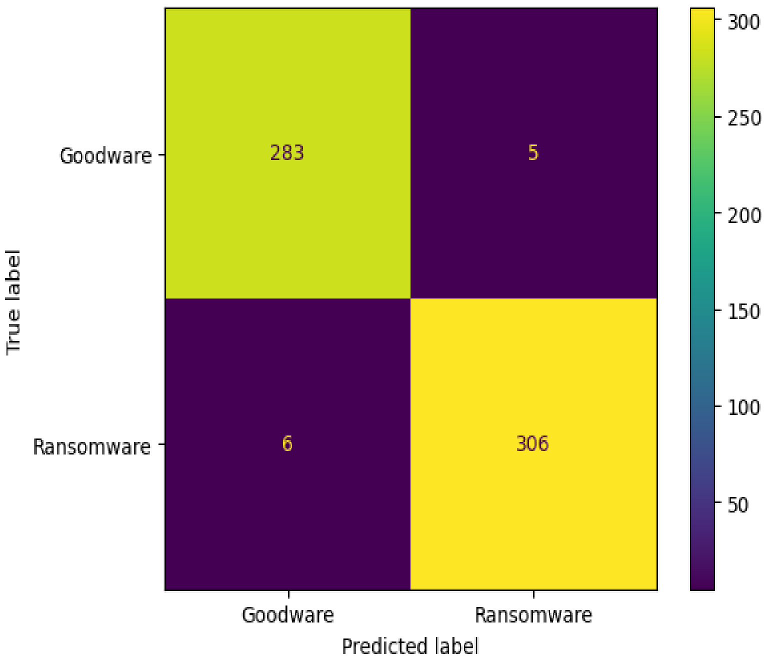


XGBoost

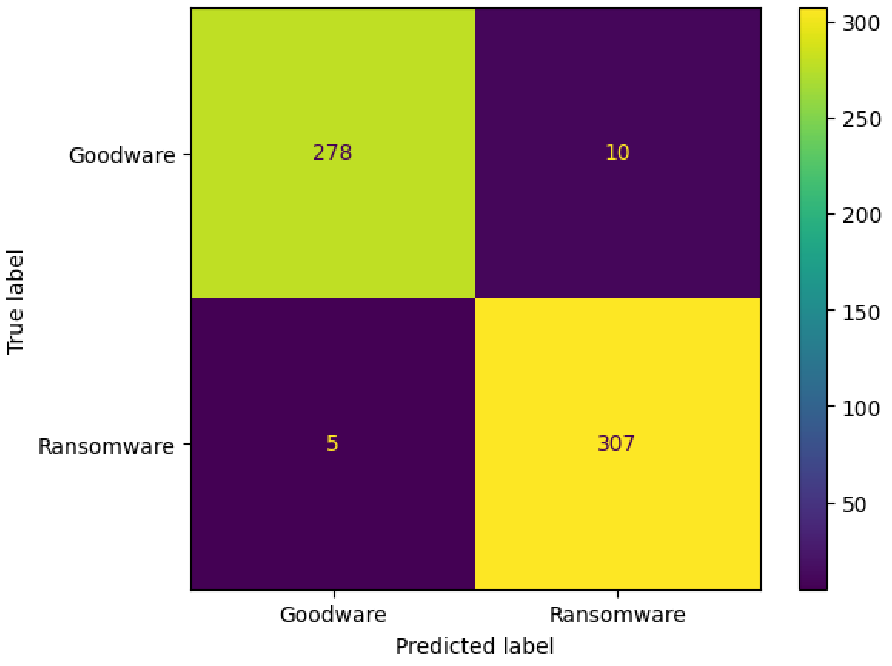


Train the Model for Dataset 1

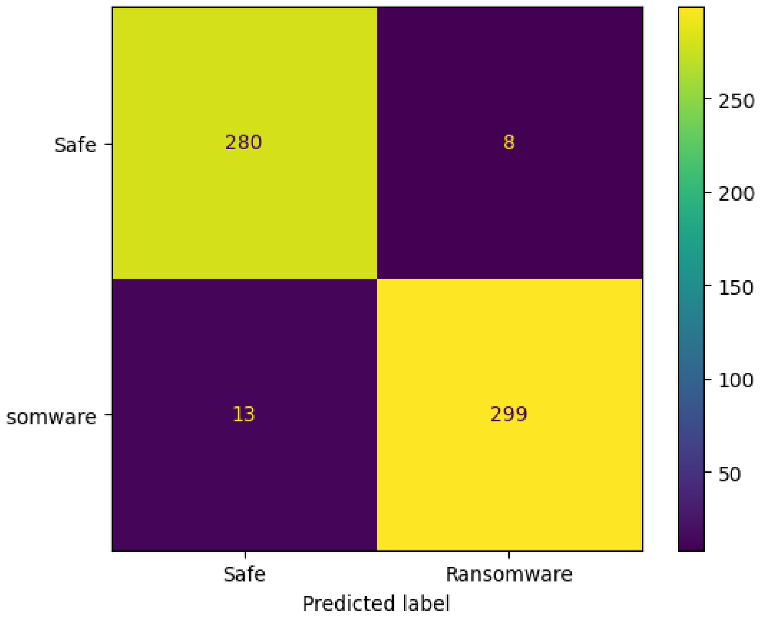
Decision Tree



KNN



Neural Network



Performance Analysis for Dataset 1

Metrics	RF	XGBoost	DT	KNN	NN		RF	NN
Accuracy	0.9967	**0.9983**	0.9817	0.9750	0.9650	Herrera et. al. [15]	**99.0**	91.92
Precision	0.9968	**0.9968**	0.9839	0.9685	0.9739		98.19	92.31
Recall	0.9968	**1.0000**	0.9808	0.9840	0.9583		96.36	90.55
F1 Score	0.9968	**0.9984**	0.9823	0.9762	0.9661		92.25	92.12
MCC	0.9933	**0.9967**	0.9633	0.9500	0.9301			
False Positive Rate	0.0035	**0.0035**	0.0174	0.0347	0.0278			
AUC Score	1.0000	**1.0000**	0.9990	0.9910	0.9971			

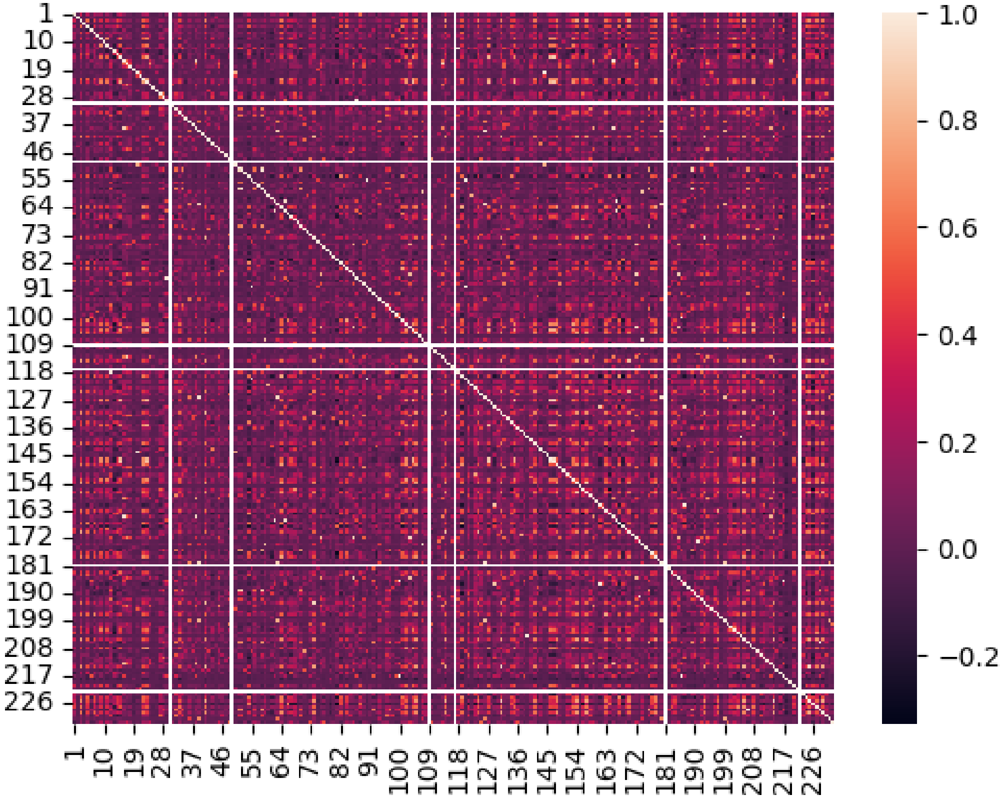
[15] Herrera-Silva, Juan A., and Myriam Hernández-Álvarez. "Dynamic feature dataset for ransomware detection using machine learning algorithms." *Sensors* 23, no. 3 (2023): 1053.

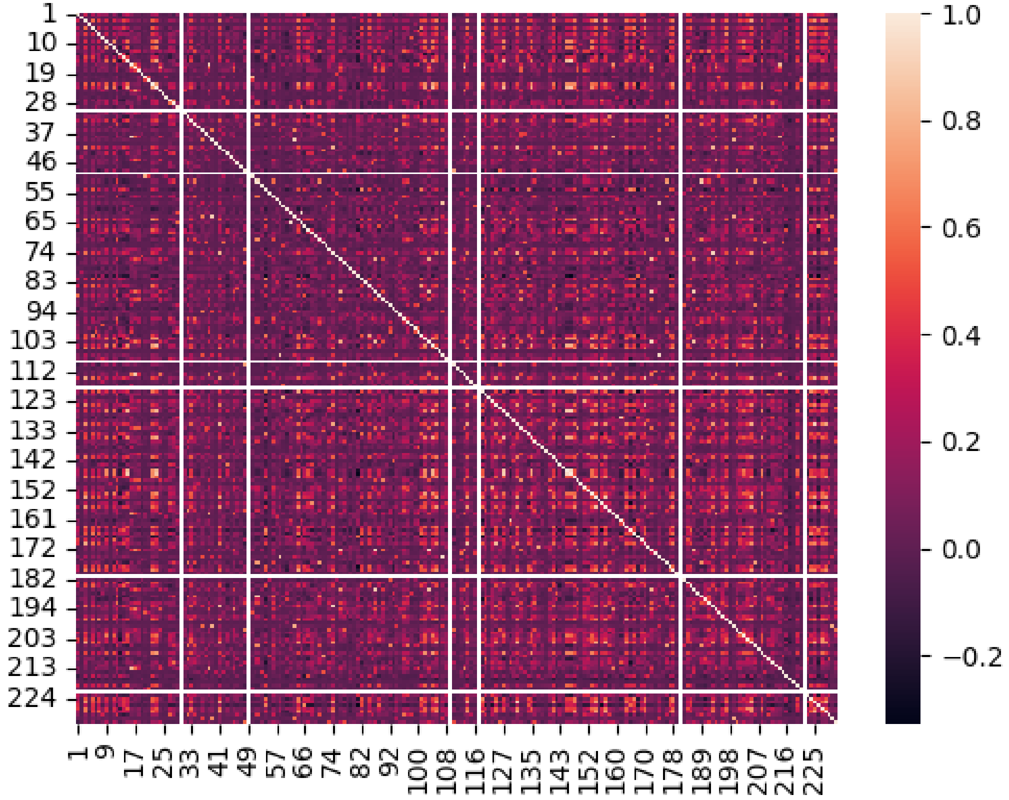
Dataset 2 Dynamic Analysis

Dropping Highly Correlated Features

- Select mostly correlated dynamic features from 232

Mathematical Model: **Pearson heatmap**





Dropped Features

['58', '87', '90', '118', '120', '126', '144', '165', '166', '180', '186', '187', '191', '204', '214', '217', '228']

Dataset 2 Dynamic Analysis

Relevant Features Extraction

- Select most relevant features and drop the less relevant features

Mathematical Model: **Information value (IV), Weight of Evidence (WoE)**

WOE = In(Distribution of Goods ÷ Distribution of Bads)

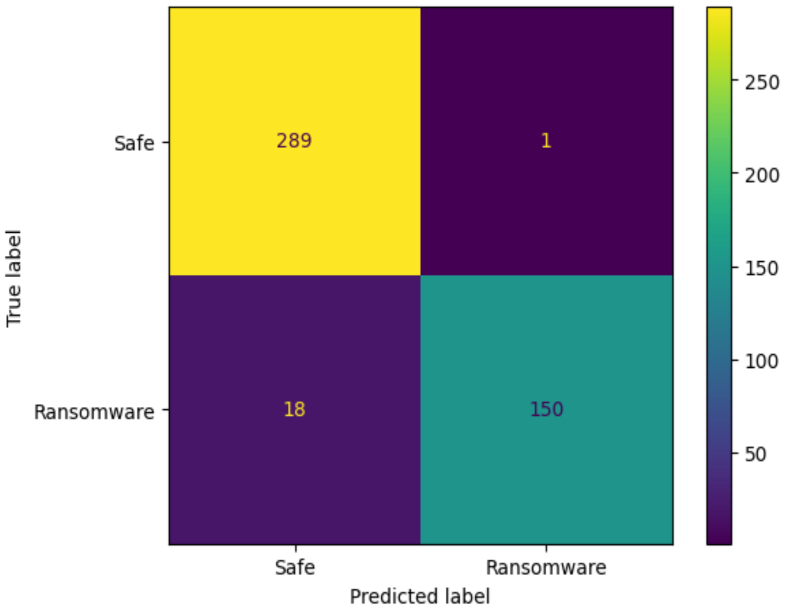
IV = ∑ (Distribution of Goods ÷ Distribution of Bads) * WOE

201 features have been selected regarding 0< IV<1

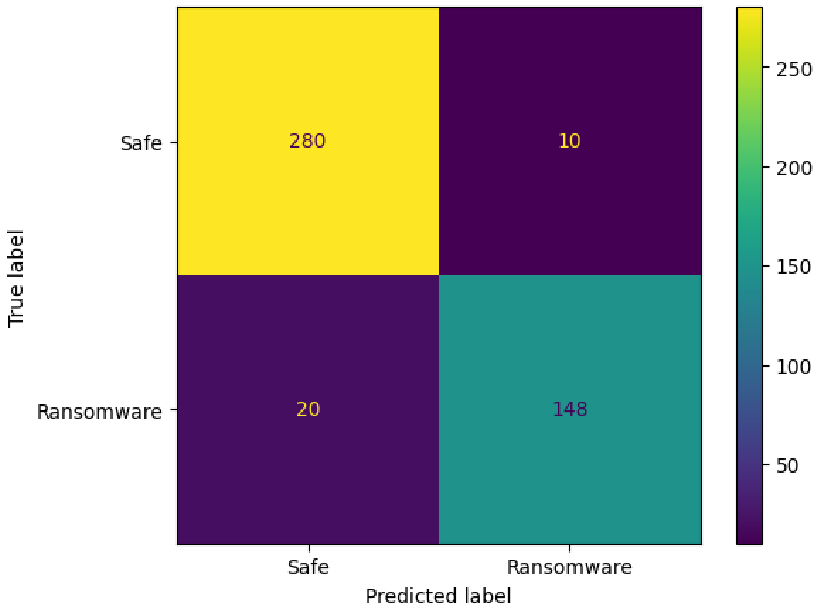
['59', '86', '174', '78', '40', '132', '44', '45', '111', '231', '196', '206', '112', '57', '42', '103', '192', '176', '21', '107', '185', '83', '193', '10', '72', '106', '219', '155', '55', '94', '133', '20', '89', '139', '47', '36', '53', '26', '80', '39', '18', '199', '173', '97', '17', '151', '158', '221', '19', '169', '136', '77', '2', '175', '137', '50', '63', '159', '172', '71', '4', '122', '198', '207', '75', '213', '116', '74', '194', '1', '35', '25', '113', '223', '209', '232', '95', '179', '14', '183', '150', '160', '37', '145', '201', '140', '60', '16', '203', '70', '128', '195', '73', '215', '152', '149', '24', '76', '32', '210', '34', '200', '38', '123', '171', '12', '184', '85', '115', '188', '93', '98', '28', '91', '189', '182', '92', '100', '125', '69', '138', '41', '29', '6', '227', '33', '130', '141', '8', '121', '110', '13', '96', '62', '31', '65', '218', '52', '51', '211', '162', '142', '56', '15', '43', '48', '108', '212', '131', '66', '79', '124', '46', '88', '208', '114', '202', '134', '164', '3', '225', '143', '205', '156', '135', '61', '67', '230', '99', '170', '190', '68', '5', '81', '9', '177', '153', '220', '216', '11', '102', '127', '104', '157', '163', '23', '105', '146', '84', '154', '147', '229', '129', '178', '161', '168', '197', '82', '27', '226', '101']

Performance Analysis for Dataset 2

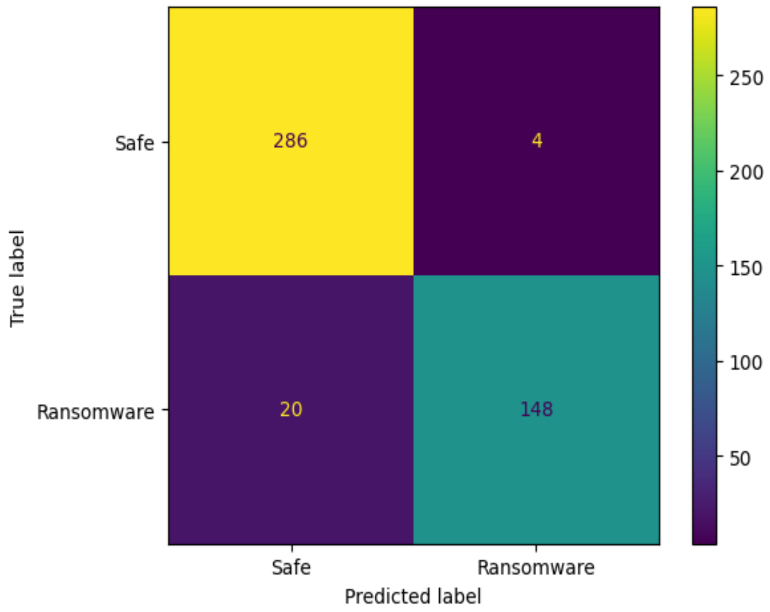
Random Forest



Decision Tree



XGBoost



Performance Analysis for Dataset 2

Random Forest

ID	Ransomware Family	RF	DT	XGBoost
1	Critroni	18	18	18
2	CryptLocker	28	27	28
3	CryptoWall	9	8	8
4	KOLLAH	9	9	9
5	Kovter	14	16	15
6	Locker	22	21	21
7	MATSNU	14	13	14
8	PGPCODER	1	1	1
9	Reveton	23	23	22
10	TeslaCrypt	2	2	2
11	Trojan-Ransom	10	10	10

Metrics	RF	DT	XGBoost
Accuracy	**0.9563**	0.9345	0.9476
Precision	**0.9868**	0.9367	0.9737
Recall	**0.8929**	0.881	0.881
F1 Score	**0.9375**	0.908	0.925
MCC	**0.9067**	0.8582	0.8875
False Positive Rate	**0.0069**	0.0345	0.0138
AUC Score	**0.9853**	0.9571	0.9892

Conclusions and Recommendations

Key Project Achievements

- ❑ Developed an effective machine-learning-based model for ransomware detection.
- ❑ Highlighted the feasibility of early ransomware detection in diverse environments.

Lessons Learned

- ➢ Importance of selecting distinctive features for enhanced accuracy.
- ➢ Challenges in balancing computational efficiency and detection performance.
- ➢ Variability in results depending on dataset quality and diversity.

Conclusions and Recommendations

Limitations

- ❑ Can’t incorporate explain ability to define the relevant features
- ❑ Potential performance degradation in real-world, unseen scenarios.
- ❑ Due to lack of large datasets zero day exploitation can’t be defended.

Future Enhancement Strategies

- ➢ Integration of real-time monitoring and automated response mechanisms.
- ➢ Exploration of hybrid detection methods combining static and dynamic features.
- ➢ Expansion to include zero-day ransomware detection using advanced techniques.
- ➢ Incorporate Explainable AI to clearly represent the relevant features to get more accuracy and less False positive rates.

Conclusions and Recommendations

Potential Real-World Applications

- Deployment in organizational cybersecurity systems to prevent data breaches.
- Utilization in antivirus and endpoint protection tools.
- Contributions to threat intelligence and proactive ransomware mitigation.

Thank You