MALLA REDDY COLLEGE OF ENGINEERING

Approved by AICTE, Permanently Affiliated to JNTUH & Accredited by NBA & NAAC

Recognized under Section 2(f) & 12(B) of the UGC Act 1956, an ISO 9001:2015 Certified Institution
Maisammaguda, Dulapally, Hyderabad-500100

DEPARTMENT OF CSE (AI & ML) AM606PC – INDUSTRIAL ORIENTED MINI PROJECT

APPROACHES FOR BENIGN AND RANSOMWARE ATTACKS DETECTION USING XGBOOST

Under the Guidance of

Mrs. Anju Gopi
Assistant Professor,
Dept. of CSE(AIML)

III-year CSE (AI & ML) - B
Team Members:

Akula Ganesh	[22Q91A6667]
Chinna Chenna Reddy Gari Sai Mokshitha	[22Q91A6678]
Palle Ramyasri	[22Q91A66B0]
Purushottam Rajpurohith	[22Q91A66B5]

Guide :
Mrs. Anju Gopi

Project Coordinator-III year
Mr. R Venkatesh

Head Of The Department
Dr. Anantha Raman G R

❖ Abstract

Ransomware attacks encrypt files and disable systems, often evading antivirus software.Traditional detection methods like process and file monitoring are resource intensive and vulnerable to manipulation by ransomware.This approach collects processor and disk I/O data from the host machine monitoring a virtual machine (VM), avoiding direct interference.

Multiple machine learning models were tested, including SVM, KNN, Decision Tree, Random Forest, XGBoost, and deep learning models like DNNand LSTM. Random Forest and XGBoost showed the best performance, achieving high accuracy and rapid detection within 400 milliseconds.The system provides real-time detection, is resilient to user workload variations, and works effectively for both known and unknown ransomware.

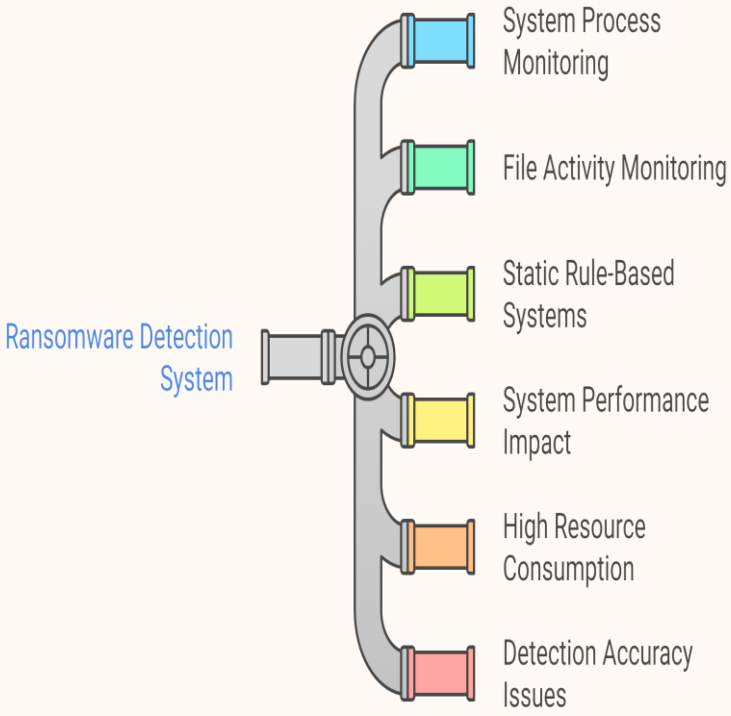
❖ Existing System

Utilizes system process monitoring to track the execution of processes for detecting ransomware.File activity monitoring observes files being created, modified, or deleted as potential indicators of malicious activity.Relies on static rule-based systems and heuristics to identify malicious behavior.The system involves continuous monitoring of all running processes, which can impact system performance.

▪ Limitations:

- **High resource consumption**, causing significant system performance degradation during monitoring.
- **Detection accuracy issues**, as newer ransomware can evade detection or interfere with the monitoring system.

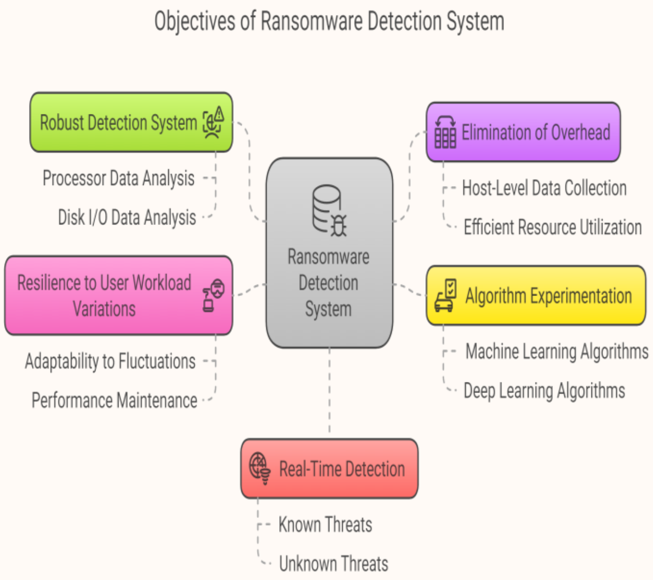
Exploring Ransomware Detection System Dynamics


Ransomware Detection System
System Process Monitoring
File Activity Monitoring
Static Rule-Based Systems
System Performance Impact
High Resource Consumption
Detection Accuracy Issues

❖ Problem Statement

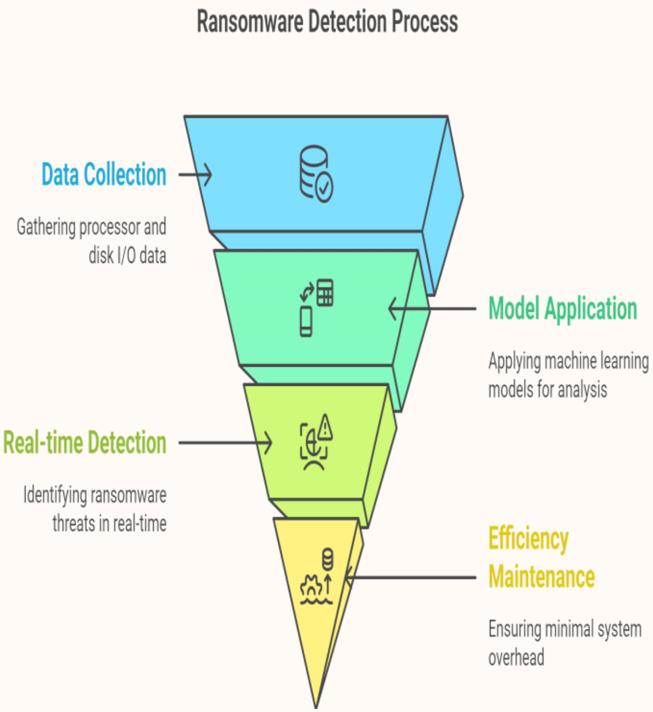
Ransomware attacks are increasingly sophisticated, often evading traditional detection methods such as antivirus software and process monitoring.Traditional systems are vulnerable to manipulation by ransomware, which may interfere with or corrupt the data being collected for detection.Current methods struggle to accurately detect unknown ransomware variants, relying heavily on predefined rules or signature-based detection.There is a need for a lightweight, real-time detection system that is resilient to changes in system workloads and can detect both known and unknown ransomware without significant performance impact.

Objective


Objectives of Ransomware Detection System
Ransomware Detection System
Robust Detection System
Processor Data Analysis
Disk I/O Data Analysis
Elimination of Overhead
Host-Level Data Collection
Efficient Resource Utilization
Resilience to User Workload Variations
Adaptability to Fluctuations
Performance Maintenance
Algorithm Experimentation
Machine Learning Algorithms
Deep Learning Algorithms
Real-Time Detection
Known Threats
Unknown Threats

To develop a robust and efficient system for detecting ransomware attacks using processor and disk I/O data.
To experiment with multiple machine learning and deep learning algorithms to identify the most accurate and fastest detection model.To build a solution that is resilient to user workload variations and can effectively detect both known and unknown ransomware in real time.

❖ Proposed System


Ransomware Detection Process
Data Collection
Gathering processor and disk I/O data
Model Application
Applying machine learning models for analysis
Real-time Detection
Identifying ransomware threats in real-time
Efficiency Maintenance
Ensuring minimal system overhead

The system collects processor and disk I/O data from the host machine running a virtual machine (VM), avoiding direct process monitoring.Uses machine learning models such as Random Forest, XGBoost, and deep learning models like DNN and LSTM for ransomware detection.The system operates with minimal overhead, making it more efficient than traditional methods that monitor every individual process.It provides real-time detection, works effectively under varying user workloads, and detects both known and unknown ransomware.

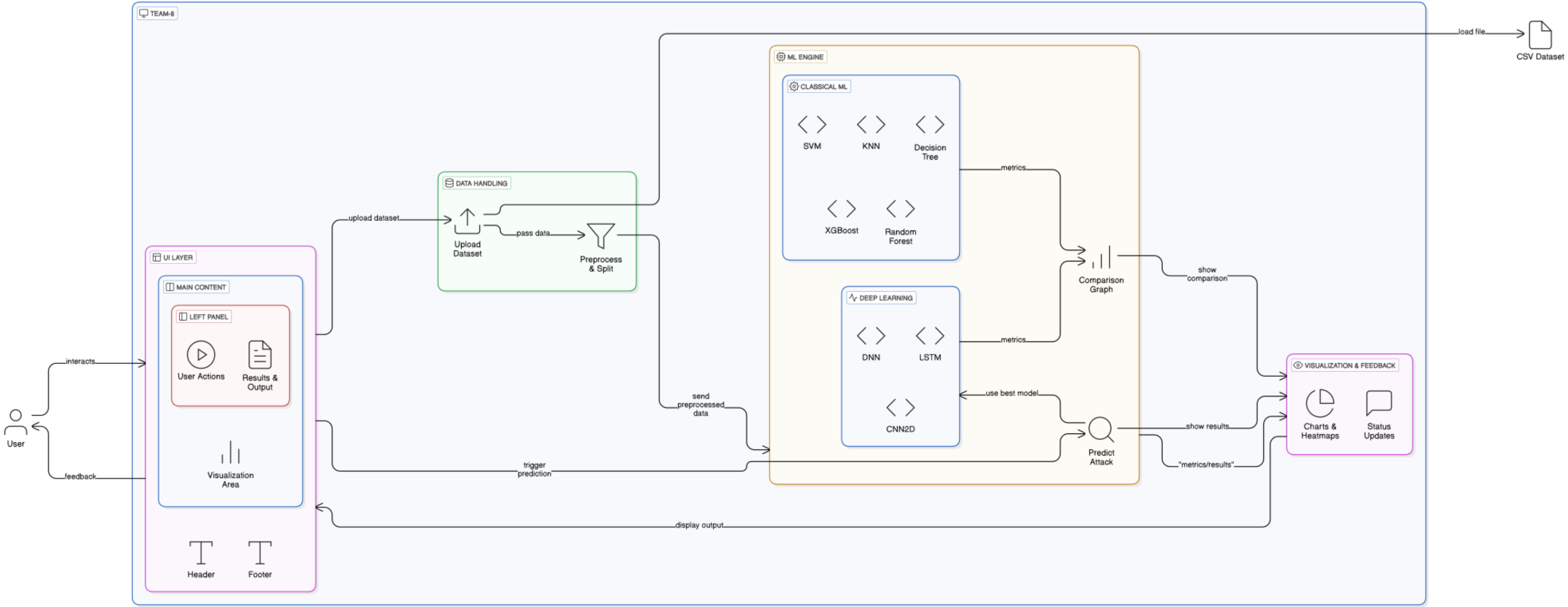
- **Advantages:**

❑ **High accuracy** with **fast detection times** (detection within 400 milliseconds) while maintaining **low system resource usage**.

❑ **Resilience** to variations in user activities, ensuring reliable detection even when user workloads change.

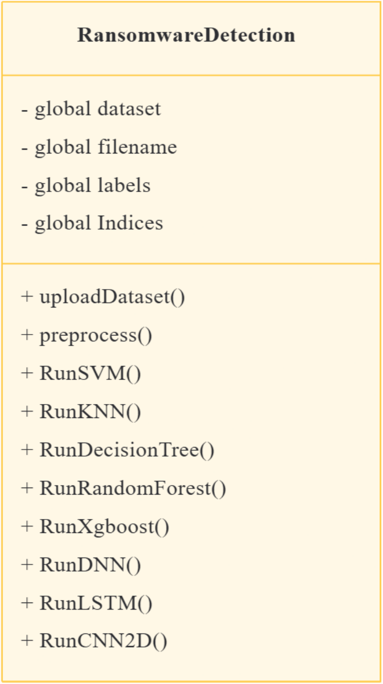
❖ System Architecture

Approaches for benign and rensomeware attacks detection using xgboost



❖ UML Diagrams

➢ Class Diagram

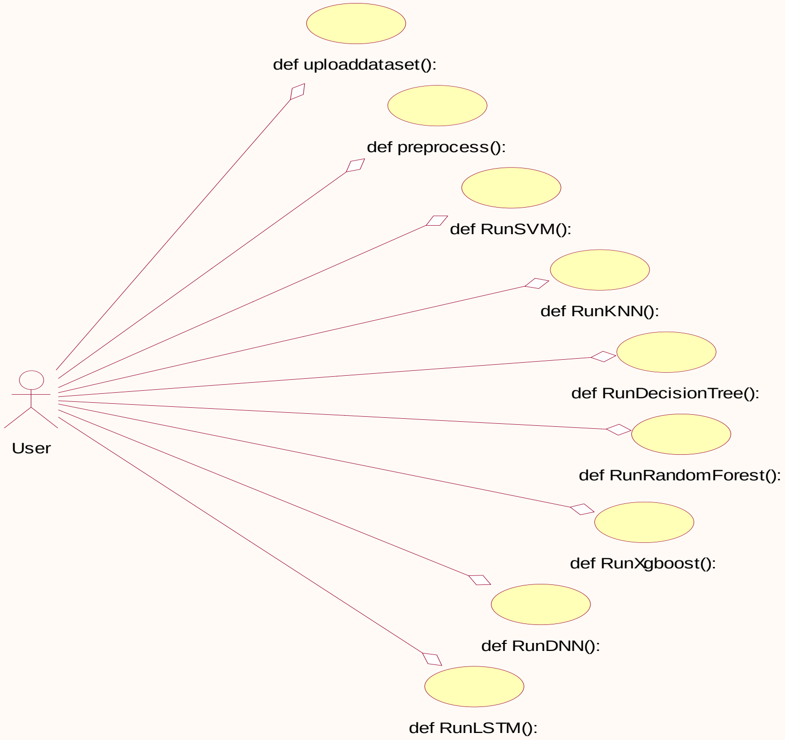


RansomwareDetection

- global dataset
- global filename
- global labels
- global Indices

+ uploadDataset()
+ preprocess()
+ RunSVM()
+ RunKNN()
+ RunDecisionTree()
+ RunRandomForest()
+ RunXgboost()
+ RunDNN()
+ RunLSTM()
+ RunCNN2D()

➢ UseCaseDiagram



User

def uploaddataset():

def preprocess():

def RunSVM():

def RunKNN():

def RunDecisionTree():

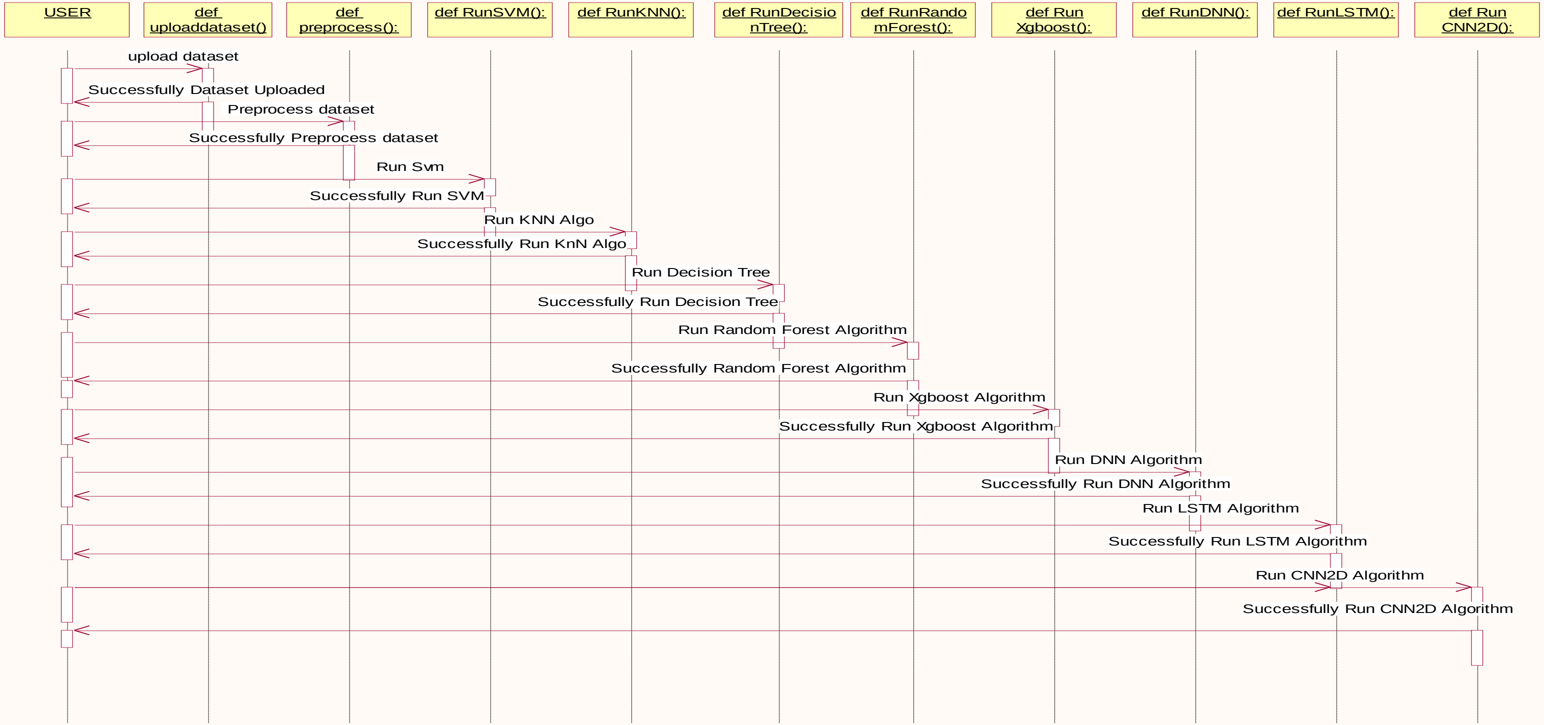
def RunRandomForest():

def RunXgboost():

def RunDNN():

def RunLSTM():

➢ Sequence Diagram



Modules Used

SVM (Support Vector Machine)

A method that finds the best boundary to separate different categories of data.

Decision Tree

A flowchart-like model that splits data based on yes/no questions to make decisions

Random Forest

A group of decision trees that vote together to make better predictions.

XGBoost (Extreme Gradient Boosting)

A fast and powerful method that builds many decision trees to fix each other's mistakes.

Matplotlib & Seaborn (Data Visualization)

Visualizes training results, confusion matrices, and performance comparisons.

Scikit-learn (Machine Learning Utilities)

Splits datasets, evaluates model performance, and calculates metrics.

DNN (Deep Neural Network)

A model with multiple layers that tries to mimic how the brain learns complex patterns.

Pandas (Data Handling)

Parses XML files containing bounding box annotations from the dataset.

LSTM (Long Short-Term Memory)

A special neural network that remembers important things over time, great for sequences like text or time series.

CNN2D (2D Convolutional Neural Network)

A type of neural network used mainly to understand images by detecting patterns like edges or shapes.

NumPy (Numerical Python)

A Python library that helps you work with large groups of numbers efficiently

FileDialog (User Interaction)

Enables users to select datasets and test images.

❖ System Requirements

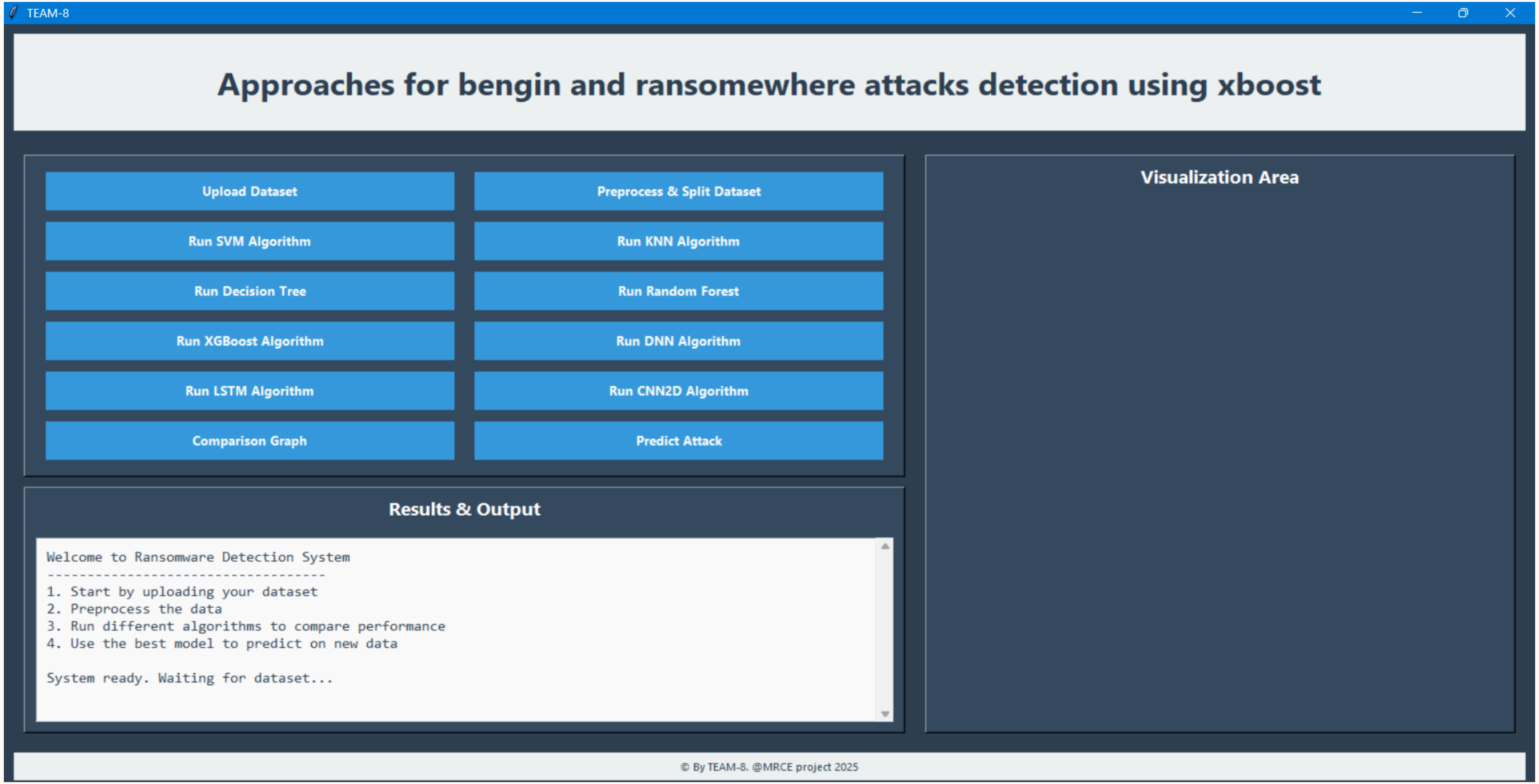
Software requirements:

Operating system	Windows 10 or 11
Frontend technologies	Tkinter,scipy,matplot-lib
Backend technologies	Tensor flow, Scikit-learn,numpy,pandas

Hardware requirements:

Processor	Intel core i3(min)
Speed	1.1Ghz
RAM	4GB (min)
Hard Disk	256GB (min)

❖ Execution Status



Approaches for bengin and ransomewhere attacks detection using xboost

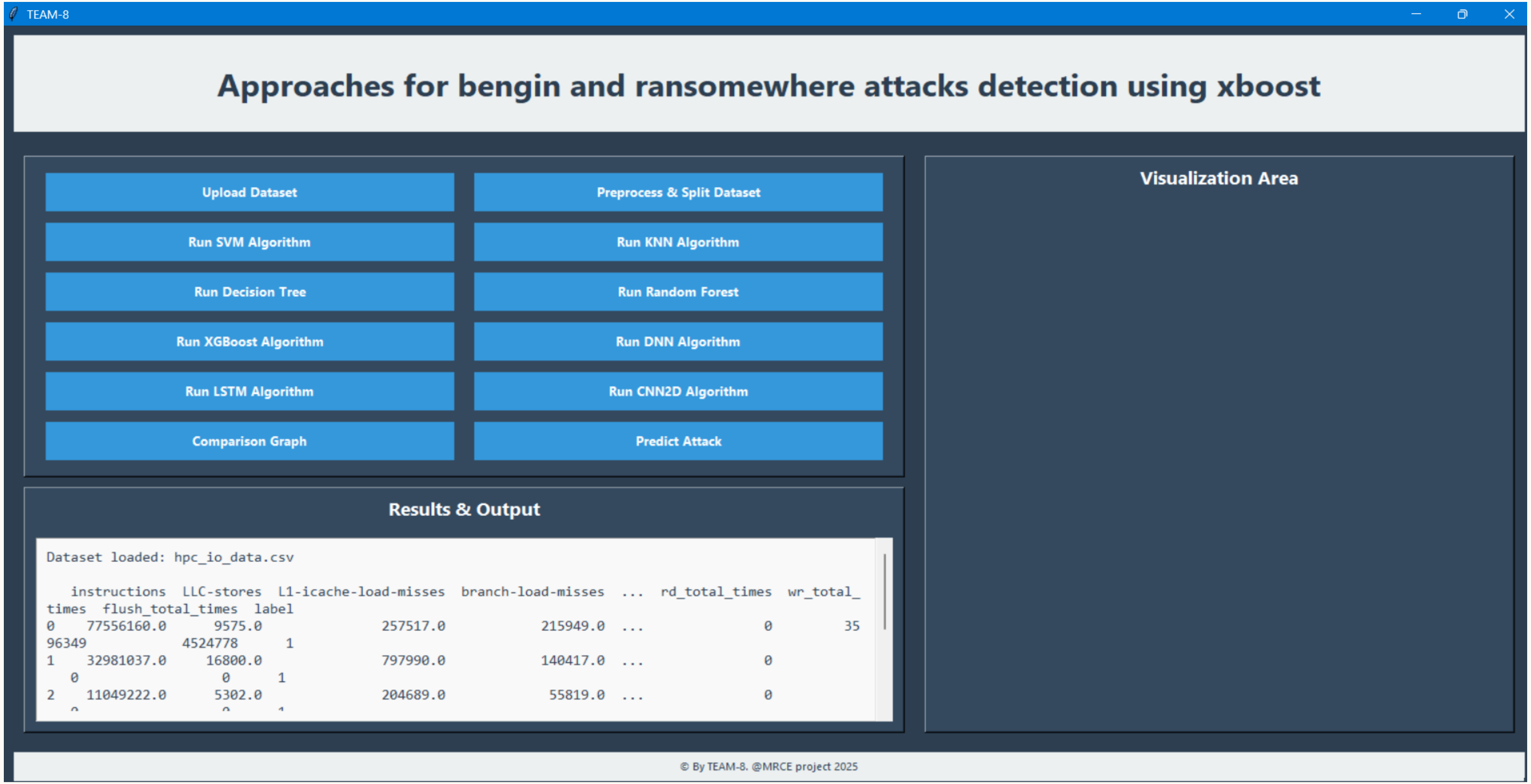
Results & Output

```
Welcome to Ransomware Detection System
-----------------------------------
1. Start by uploading your dataset
2. Preprocess the data
3. Run different algorithms to compare performance
4. Use the best model to predict on new data

System ready. Waiting for dataset...
```

© By TEAM-8. @MRCE project 2025

➤ Dataset uploading



Approaches for bengin and ransomewhere attacks detection using xboost

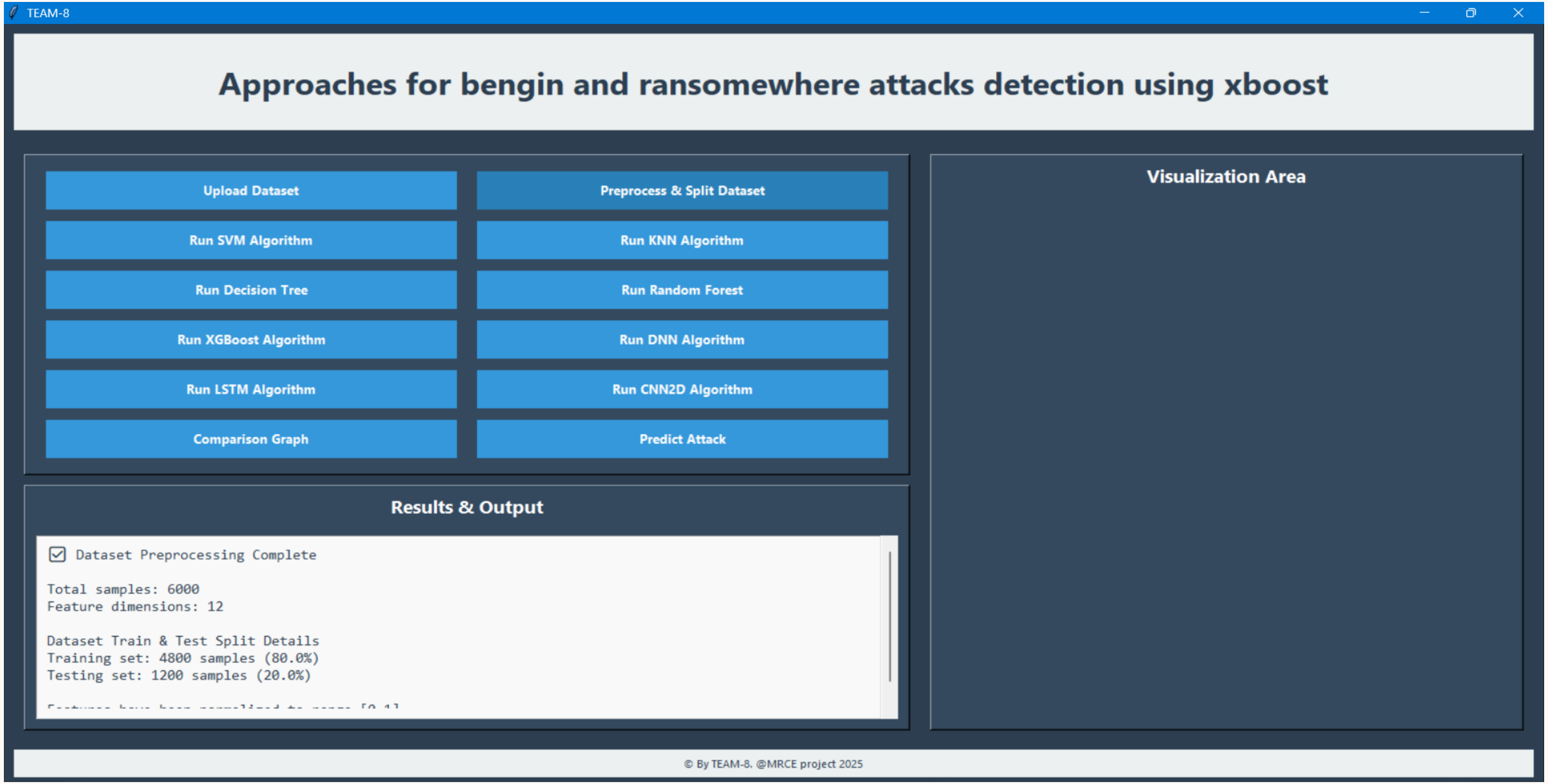
Results & Output

```
Dataset loaded: hpc_io_data.csv

   instructions  LLC-stores  L1-icache-load-misses  branch-load-misses  ...  rd_total_times  wr_total_
times  flush_total_times  label
0    77556160.0      9575.0               257517.0            215949.0  ...               0          35
96349              4524778      1
1    32981037.0     16800.0               797990.0            140417.0  ...               0
   0                  0      1
2    11049222.0      5302.0               204689.0             55819.0  ...               0
```

© By TEAM-8. @MRCE project 2025

➢ preprocessing



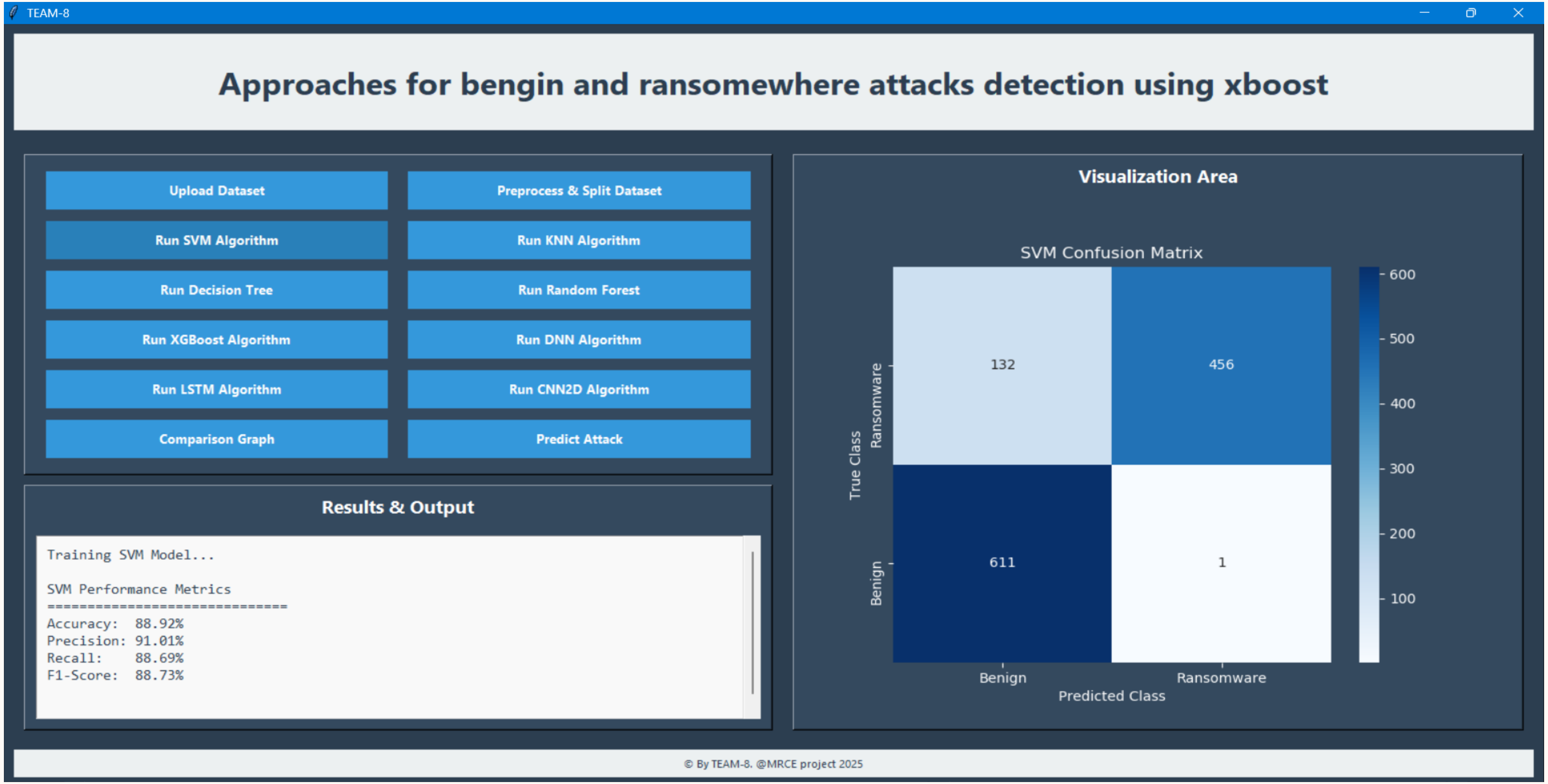
```
☑ Dataset Preprocessing Complete

Total samples: 6000
Feature dimensions: 12

Dataset Train & Test Split Details
Training set: 4800 samples (80.0%)
Testing set: 1200 samples (20.0%)

Features have been normalized to range [0,1]
```

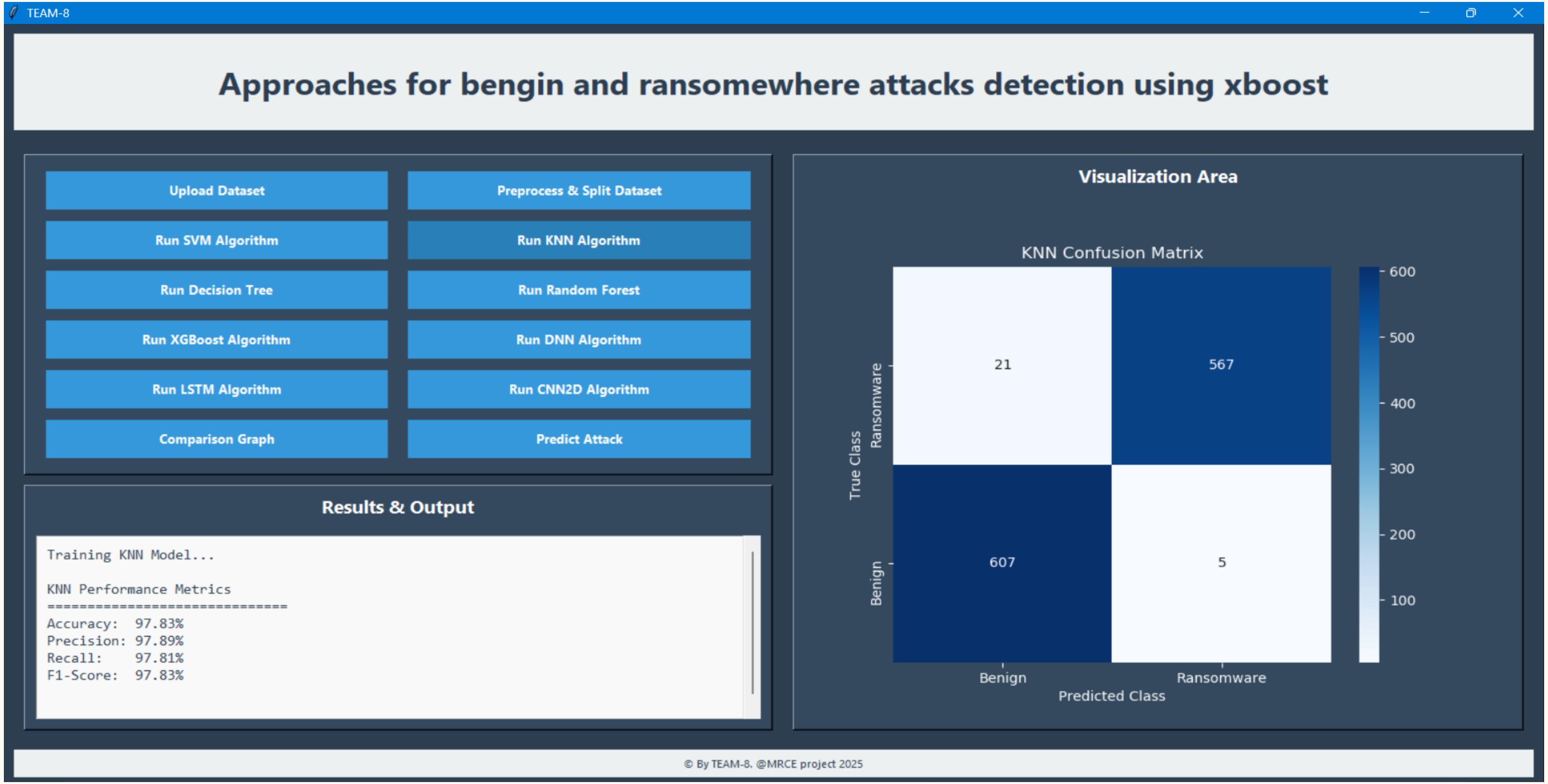
➢ SVM algorithm



```
Training SVM Model...

SVM Performance Metrics
==============================
Accuracy:   88.92%
Precision:  91.01%
Recall:     88.69%
F1-Score:   88.73%
```

➢ KNN algorithm



Approaches for bengin and ransomewhere attacks detection using xboost

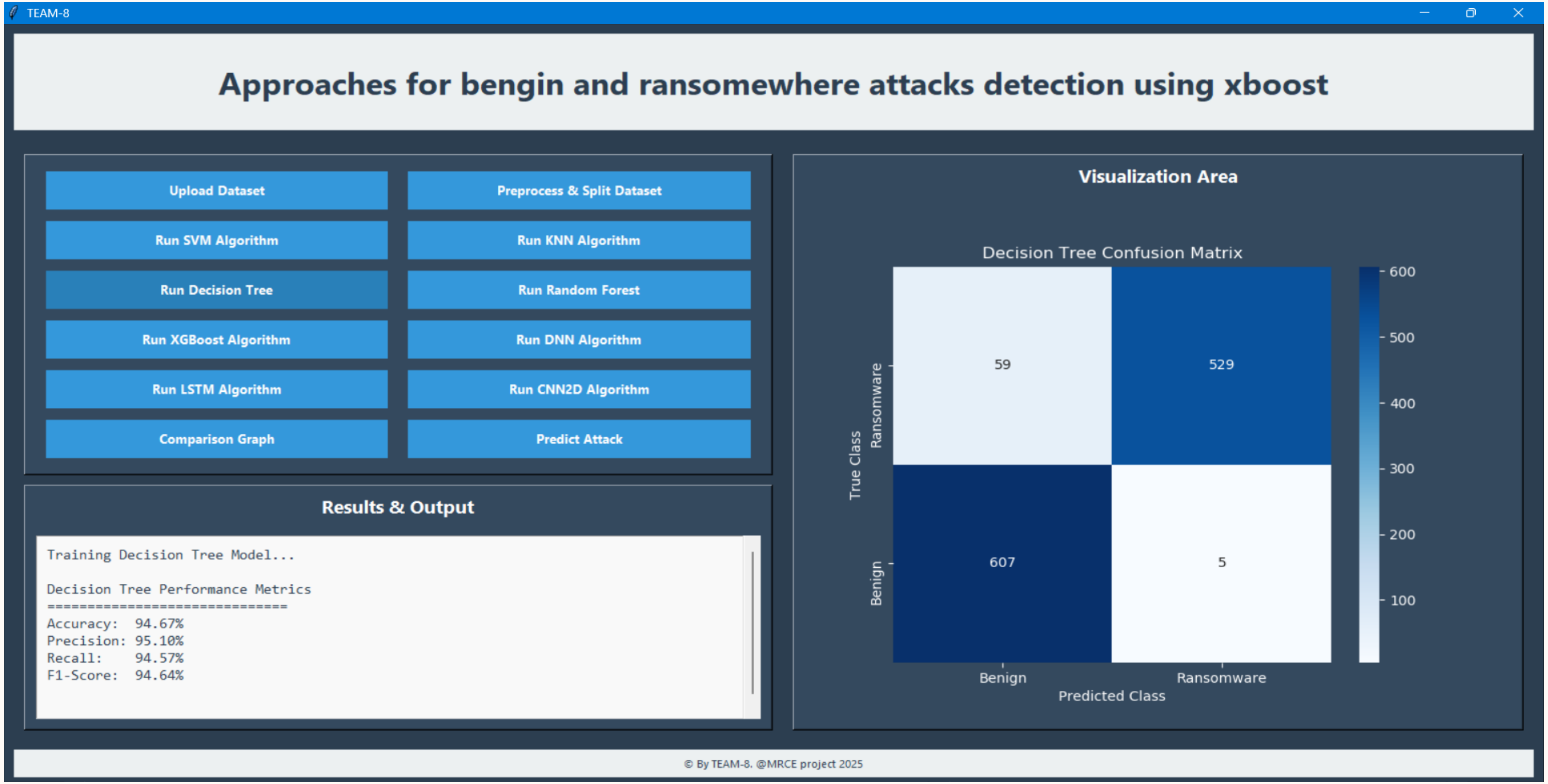
Results & Output

```
Training KNN Model...

KNN Performance Metrics
==============================
Accuracy:  97.83%
Precision: 97.89%
Recall:    97.81%
F1-Score:  97.83%
```

© By TEAM-8. @MRCE project 2025

➤ Decision tree algorithm



Approaches for bengin and ransomewhere attacks detection using xboost

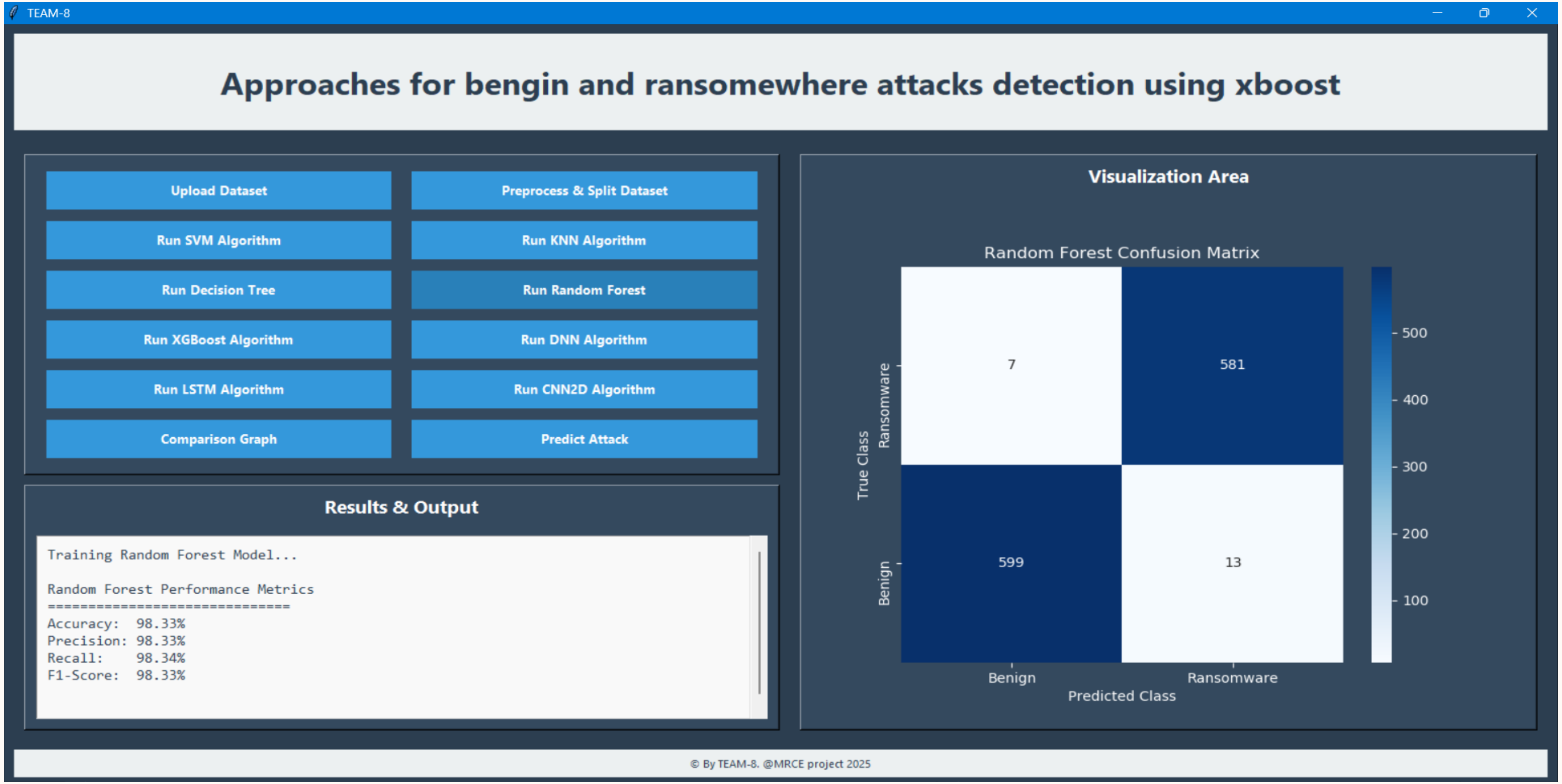
Results & Output

```
Training Decision Tree Model...

Decision Tree Performance Metrics
==============================
Accuracy:   94.67%
Precision:  95.10%
Recall:     94.57%
F1-Score:   94.64%
```

© By TEAM-8. @MRCE project 2025

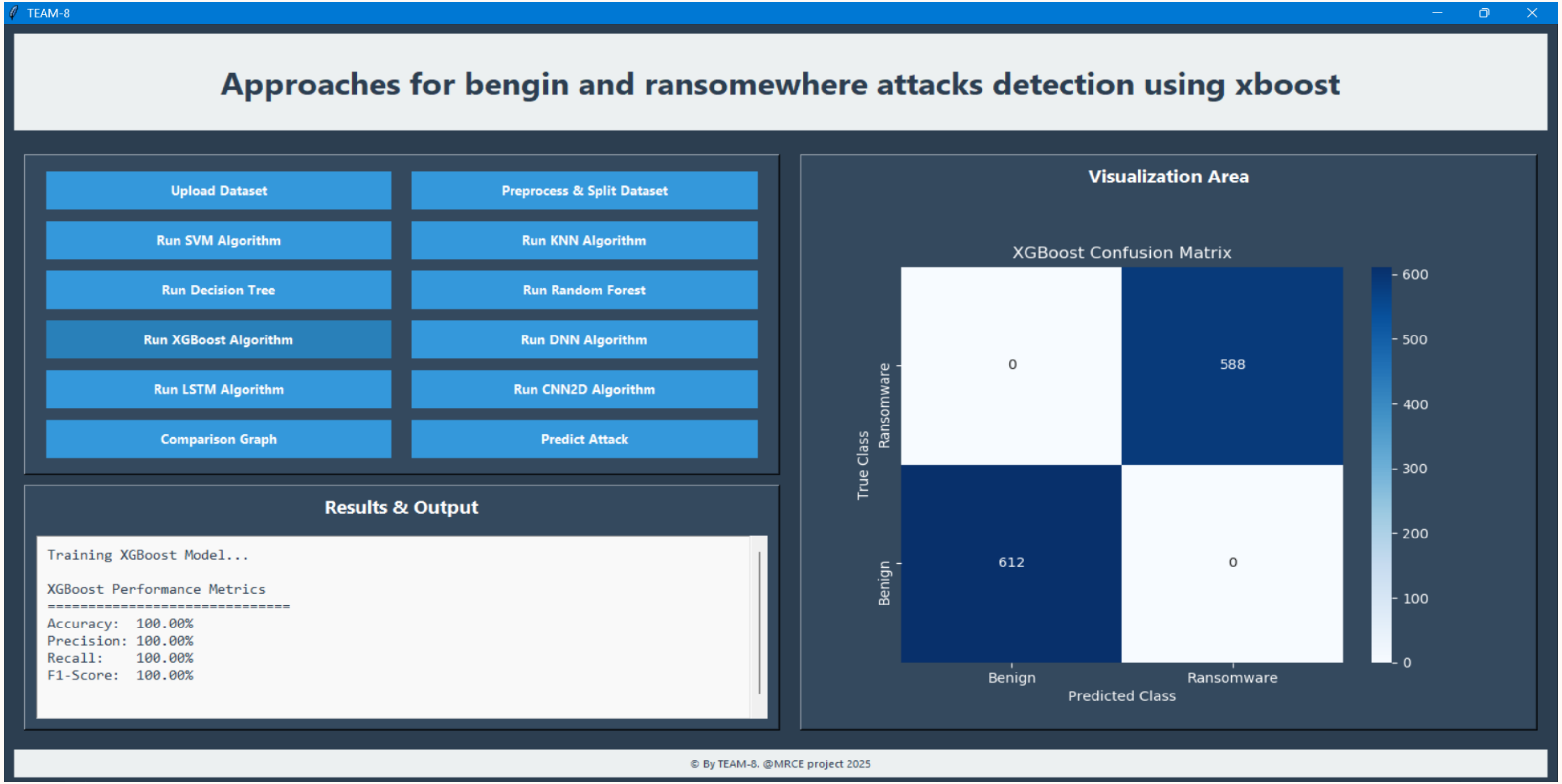
➢ Random Forest algorithm



```
Training Random Forest Model...

Random Forest Performance Metrics
==============================
Accuracy:  98.33%
Precision: 98.33%
Recall:    98.34%
F1-Score:  98.33%
```

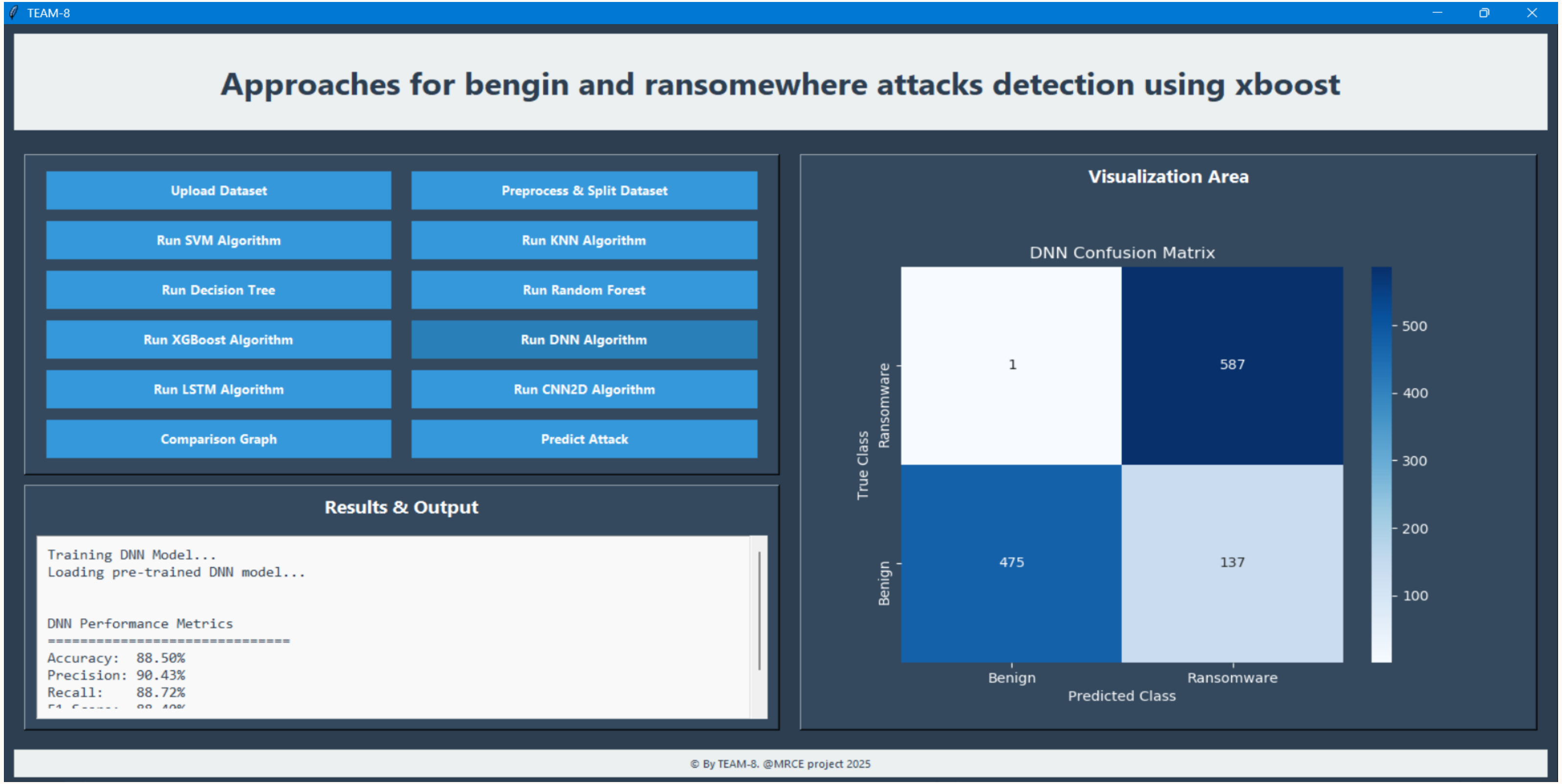
XGBoost algorithm



```
Training XGBoost Model...

XGBoost Performance Metrics
==============================
Accuracy:  100.00%
Precision: 100.00%
Recall:    100.00%
F1-Score:  100.00%
```

➢ DNN algorithm



Approaches for bengin and ransomewhere attacks detection using xboost

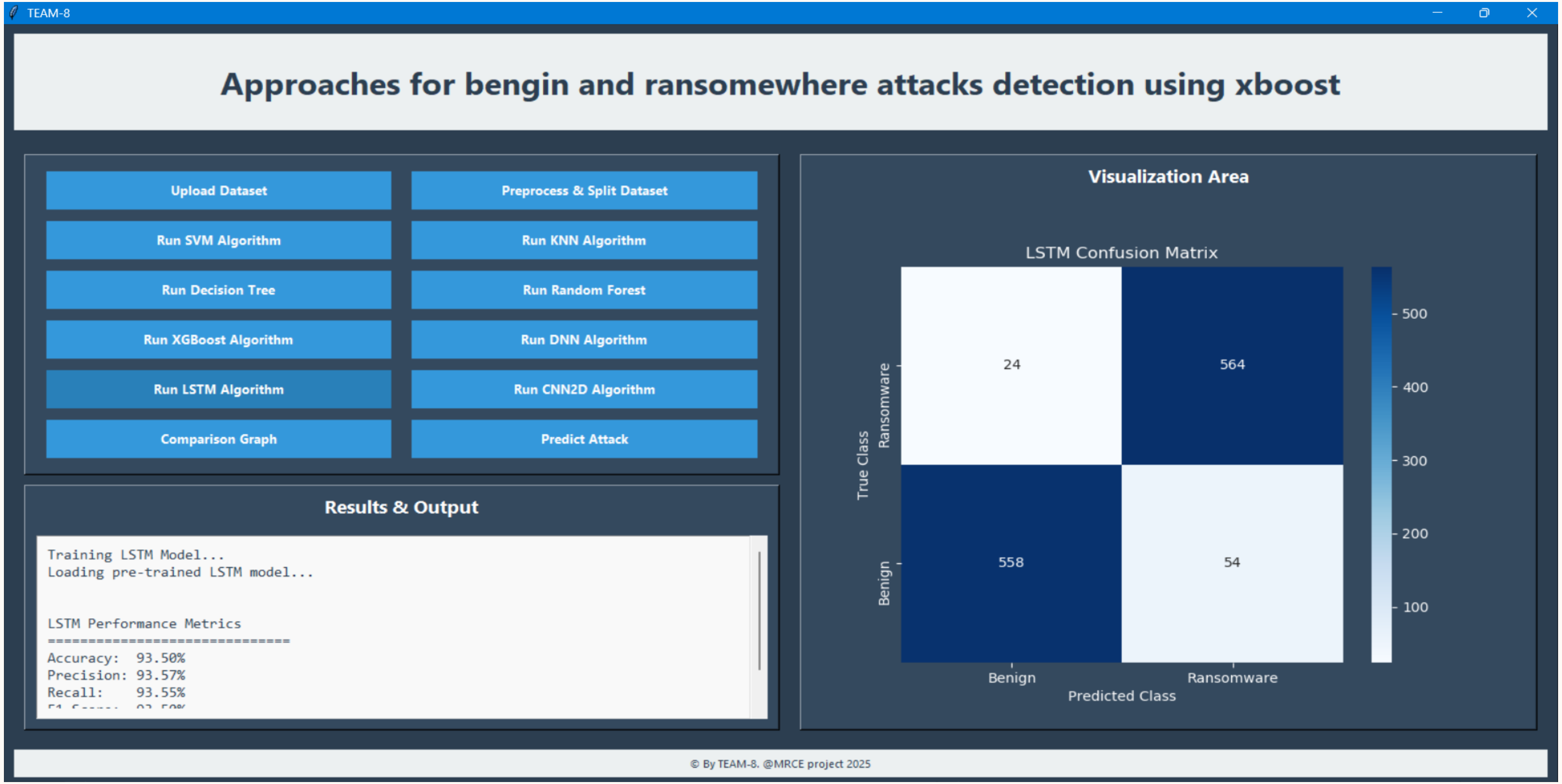
Results & Output

```
Training DNN Model...
Loading pre-trained DNN model...


DNN Performance Metrics
==============================
Accuracy:  88.50%
Precision: 90.43%
Recall:    88.72%
F1 Score:  88.40%
```

© By TEAM-8. @MRCE project 2025

LSTM algorithm



Approaches for bengin and ransomewhere attacks detection using xboost

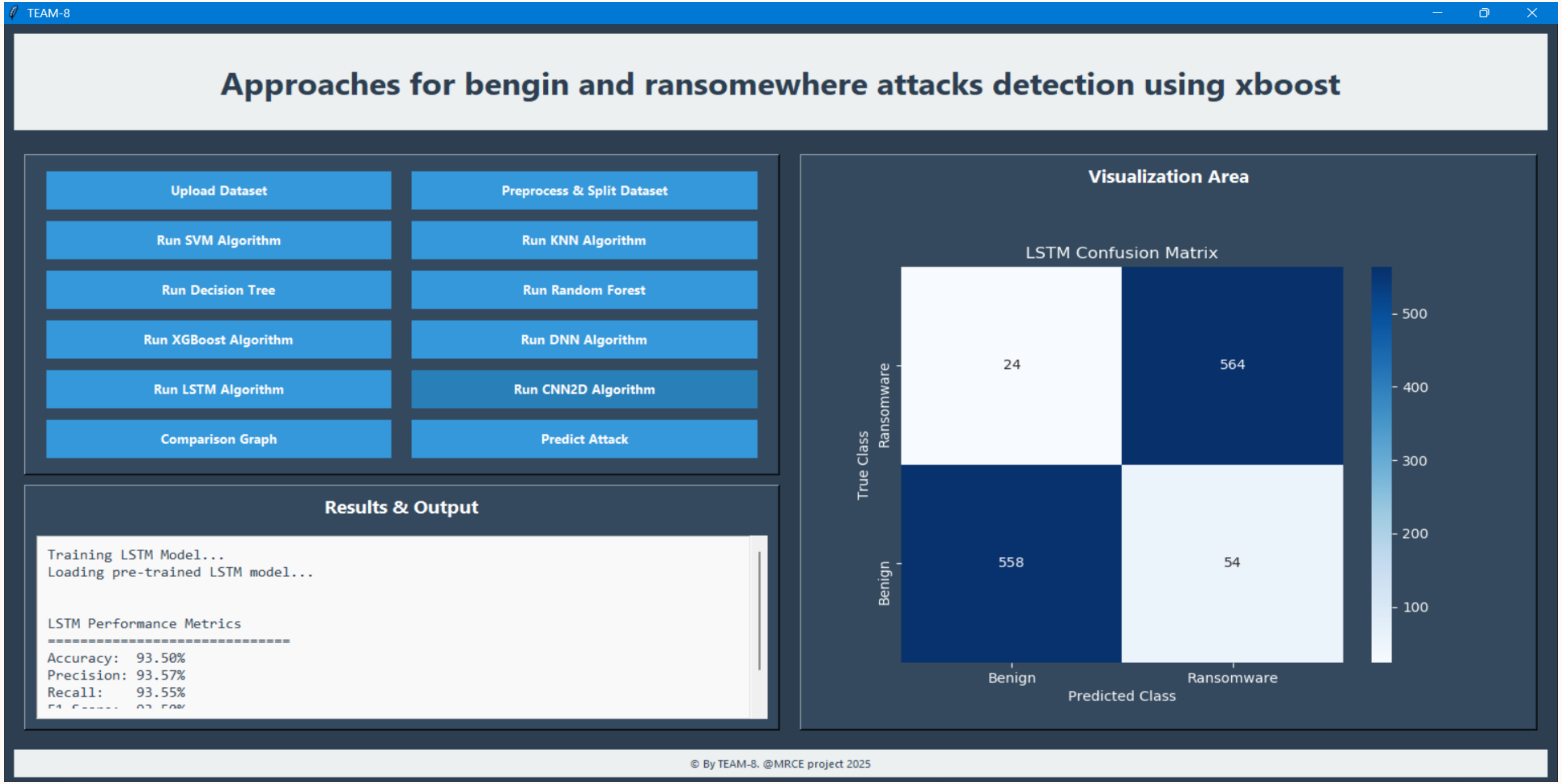
Results & Output

```
Training LSTM Model...
Loading pre-trained LSTM model...


LSTM Performance Metrics
==============================
Accuracy:  93.50%
Precision: 93.57%
Recall:    93.55%
F1 Score:  93.50%
```

© By TEAM-8. @MRCE project 2025

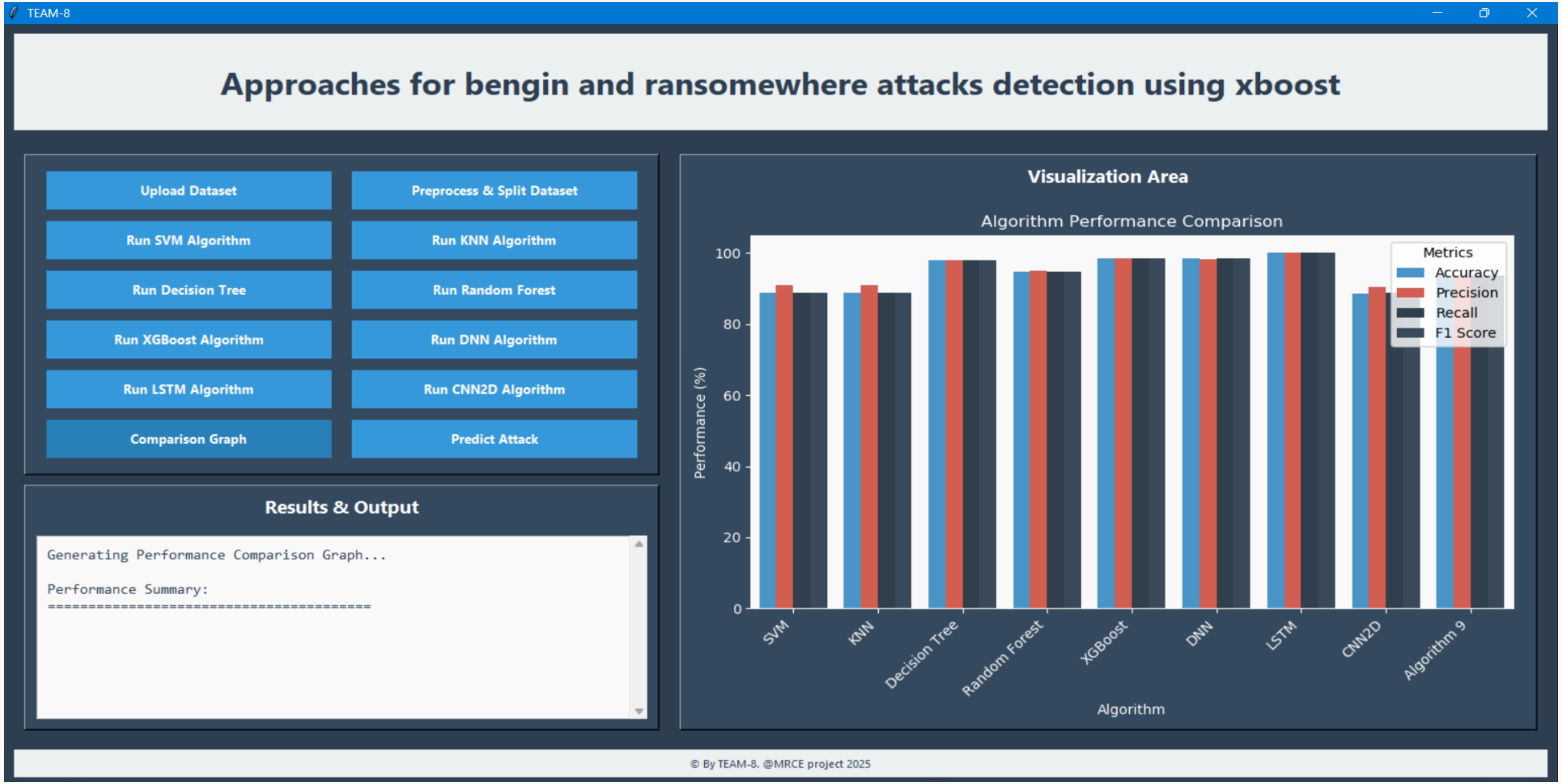
➢ CNN2D algorithm



```
Training LSTM Model...
Loading pre-trained LSTM model...


LSTM Performance Metrics
==============================
Accuracy:  93.50%
Precision: 93.57%
Recall:    93.55%
F1 Score:  93.50%
```

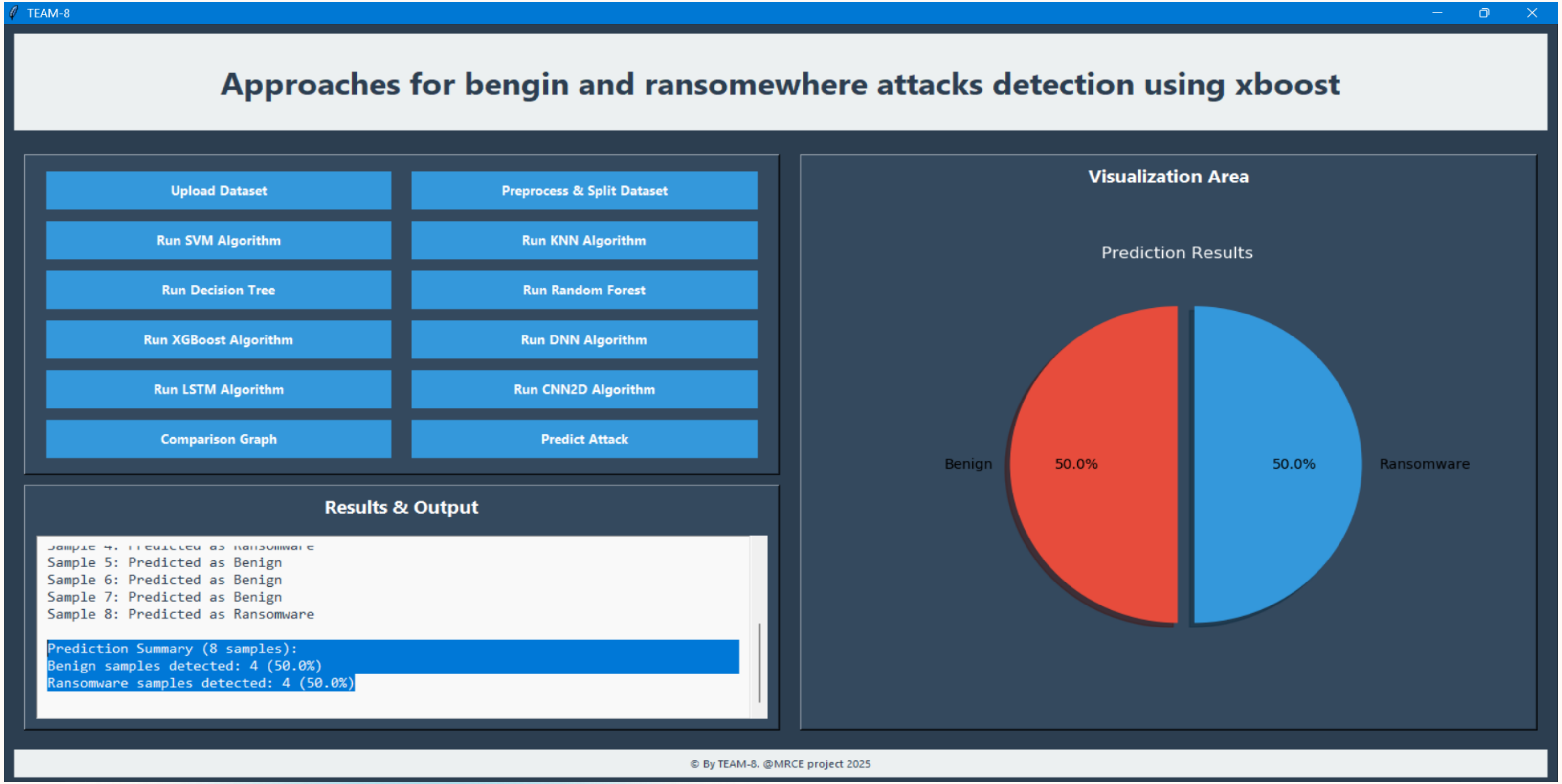
➢ Comparison graph



```
Generating Performance Comparison Graph...

Performance Summary:
========================================
```

➢ Attack prediction



```
Sample 4: Predicted as Ransomware
Sample 5: Predicted as Benign
Sample 6: Predicted as Benign
Sample 7: Predicted as Benign
Sample 8: Predicted as Ransomware

Prediction Summary (8 samples):
Benign samples detected: 4 (50.0%)
Ransomware samples detected: 4 (50.0%)
```

References

- [1] SR Department. (2025). Ransomware victimization rate 2022. Accessed: Apr. 6, 2022. [Online]. Available: **https://www.statista. com/statistics/204457/businesses-ransomware-attack-rate/**
- [2] D. Braue. (2025). Ransomware Damage Costs. Accessed: Sep. 16, 2022. [Online]. Available**: https://cybersecurityventures.com/globalransomware-damage-costs-predicted-to-reach-250-billion-usd-by-2031/**
- [3] Logix Consulting. (2025). What is Signature Based Malware Detection. Accessed: Apr. 3, 2023. [Online]. Available**: https://www.logixconsulting. com/2020/12/15/what-is-signature-based-malware-detection/**
- [4] W. Liu, P. Ren, K. Liu, and H.-X. Duan(2024), ''Behavior-based malware analysis and detection,'' in Proc. 1st Int. Workshop Complex. Data Mining, Sep. 2011, pp. 39–42.
- [5] (2024). Polymorphic Malware. Accessed: Apr. 3, 2023. [Online]. Available: **https://www.thesslstore.com/blog/polymorphic-malware-andmetamorphic-malware-what-you-need-to-know/**
- [6] M. Loman. (2024). Lockfile Ransomware's Box of Tricks: Intermittent Encryption and Evasion. Accessed: Nov. 16, 2021. [Online]. Available: **https://news.sophos.com/en-us/2021/08/27/lockfile-ransomwares-box-oftricks-intermittent-encryption-and-evasion/**

Thank you!

Do you have any queries!

By- team-8 (csm-b 3rd)