AI-Powered Insider Threat Detection

Group Members:

21K-3061
21K-3062
21K-3006

Abstract

This project presents an AI-powered insider threat detection system utilizing the CICIDS 2018 dataset. The system leverages machine learning techniques, specifically Random Forest and Convolutional Neural Network (CNN) models, to classify network traffic into benign or various attack types. The approach includes comprehensive data preprocessing, exploratory data analysis, model training, evaluation, and visualization to develop an effective intrusion detection system. The results demonstrate promising accuracy and robustness, highlighting the potential of AI in cybersecurity threat detection.

Introduction

Insider threats pose significant risks to organizational cybersecurity, often leading to data breaches and system compromises. Detecting such threats requires sophisticated methods capable of analyzing complex network traffic patterns. The CICIDS 2018 dataset provides a comprehensive collection of network traffic data labeled with various attack types, making it an ideal benchmark for developing intrusion detection systems. This project implements an AI-based detection system using machine learning models to classify network traffic and identify potential insider threats accurately.

Literature Review

Several studies have explored machine learning for intrusion detection, focusing on algorithms such as Random Forest, Support Vector Machines, and deep learning models including CNNs. Random Forest classifiers are known for their robustness and interpretability, providing feature importance insights. CNNs, traditionally used in image processing, have shown effectiveness in capturing spatial and temporal patterns in sequential data like network traffic. Prior work emphasizes the importance of data preprocessing, balancing datasets, and feature normalization to enhance model performance. The CICIDS 2018 dataset has been widely adopted in research for benchmarking intrusion detection models due to its diversity and realism.

Methodology

Data Loading and Exploratory Data Analysis (EDA)

The CICIDS 2018 dataset was loaded using pandas, and initial EDA was performed to understand data characteristics. This included checking dataset shape, missing values, attack distribution, data types, and feature correlations. Visualizations such as count plots for attack types, correlation heatmaps, and interactive scatter plots were generated to analyze feature relationships and class imbalances.

Data Preprocessing

Data preprocessing involved:

- Removing missing values.
- Encoding categorical attack labels into numerical format using LabelEncoder.
- Balancing the dataset by resampling each attack class to a uniform sample size to mitigate class imbalance.
- Dropping non-numeric and identifier columns.
- Handling infinite and extreme values by replacing infinities with medians and clipping outliers based on interquartile ranges.
- Normalizing features with MinMaxScaler.
- Splitting data into training and testing sets with stratification to preserve class distribution.
- Preparing data for CNN input by reshaping features to 3D tensors.
- Saving preprocessing artifacts such as label encoders, scalers, and feature names for reproducibility.

Model Development

Two models were developed and trained:

1. **Random Forest Classifier**
    - Configured with 100 trees, max depth of 20, and minimum samples split and leaf parameters to prevent overfitting.
    - Trained on the balanced and normalized feature set.
    - Performance evaluated using accuracy, confusion matrix, and classification report.
    - Feature importance was analyzed and visualized to identify key predictors.
    - The trained model was saved for future inference.
2. **Convolutional Neural Network (CNN)**
    - Designed with three Conv1D layers with increasing filter sizes (64, 128, 256), each followed by batch normalization and max pooling.
    - Fully connected dense layers with dropout were added to reduce overfitting.
    - The model was compiled with categorical crossentropy loss and Adam optimizer.
    - Trained for up to 30 epochs with early stopping and model checkpoint callbacks based on validation accuracy.
    - Training history was plotted to visualize accuracy and loss trends.
    - The best model was saved for deployment.

Conclusion

The project successfully implemented an AI-powered insider threat detection system using the CICIDS 2018 dataset. Both Random Forest and CNN models demonstrated strong classification performance, with the CNN providing a deep learning approach capable of capturing complex feature interactions. The comprehensive preprocessing pipeline ensured balanced and clean data, contributing to model effectiveness. This work confirms the viability of AI techniques in enhancing cybersecurity defenses against insider threats and lays the groundwork for further improvements such as real-time deployment and integration with network monitoring tools.

References

- CICIDS 2018 Dataset. Canadian Institute for Cybersecurity.
- Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
- Chollet, F. (2017). Deep Learning with Python. Manning Publications.
- Scikit-learn Documentation. https://scikit-learn.org
- TensorFlow Keras Documentation. https://www.tensorflow.org/api_docs/python/tf/keras

Citations:

1. https://ppl-ai-file-upload.s3.amazonaws.com/web/direct-files/attachments/62031108/faa0ee32-debb-44c8-b008-21e9ab2c4da1/paste.txt