ALY6980 - CAPSTONE

MID-TERM PRESENTATION

MODULE 7 ASSIGNMENT

By:

Ansh Aya

Suprit Mestry

Venkat Rishi Kumar Aloor

Kelvina Pethani

Lalitesh

To:

Prof: Valerie Atherley

Executive Summary

- This document will examine the methods used, objectives sought, and key advantages of using ML/AI for software security.
- Identifying vulnerabilities more accurately and efficiently.
- Prioritizing vulnerabilities before they can be exploited.
- Automating security tasks.

AIM OF THE PROJECT

- Automate vulnerability assessment to reduce manual labor and improve efficiency
- Realize cost savings by minimizing human resource dependence
- Enable rapid responses to evolving security threats by quickly identifying and addressing vulnerabilities
- Improve overall security while optimizing resource allocation.

BUSINESS QUESTIONS

- How can we use machine learning and artificial intelligence to collect and evaluate threat intelligence data so that we can proactively fight against new threats?
- Is it possible to utilize ML/AI to improve email security by spotting phishing emails and harmful attachments?
- How might ML/AI help automate and speed up incident response, such as determining the extent of an invasion of privacy and containing it?
- How can ML/AI improve the accuracy of our security alerts so that our safety personnel can concentrate on actual threats?
- How can we use ML/AI to evaluate the safety record of suppliers and third-party vendors?

TECHNIQUES & METHODOLOGY USED

- Data Acquisition: Collect datasets from diverse sources.
- Data Exploration: Visualize data, identify patterns, and relationships.
- Data Refinement: Cleanse and preprocess data for quality.
- Machine Learning and AI: Analyze large datasets efficiently.
- Data Modeling: Structure and manipulate data for AI.
- Additional Techniques: Understand data, clean, transform, filter, summarize, and perform feature engineering.
- Initial Analysis: Study the product and its dataset for understanding vulnerabilities.
- ML and AI Model: Create a model for automated vulnerability assessment, learning from past data to enhance accuracy.

Data Analysis - Descriptive Statistics

Descriptive Statistics: Calculating measures like mean, median, mode, standard deviation, variance, and quartiles to summarize the dataset's central tendency, dispersion, and shape.

Data Visualization: Creating visual representations like histograms, scatter plots, box plots, heat maps, etc., to understand distributions, correlations, and outliers.

Exploratory Data Analysis (EDA): This involves a thorough examination of the dataset, identifying patterns and anomalies, and understanding relationships using statistical and visualization tools.

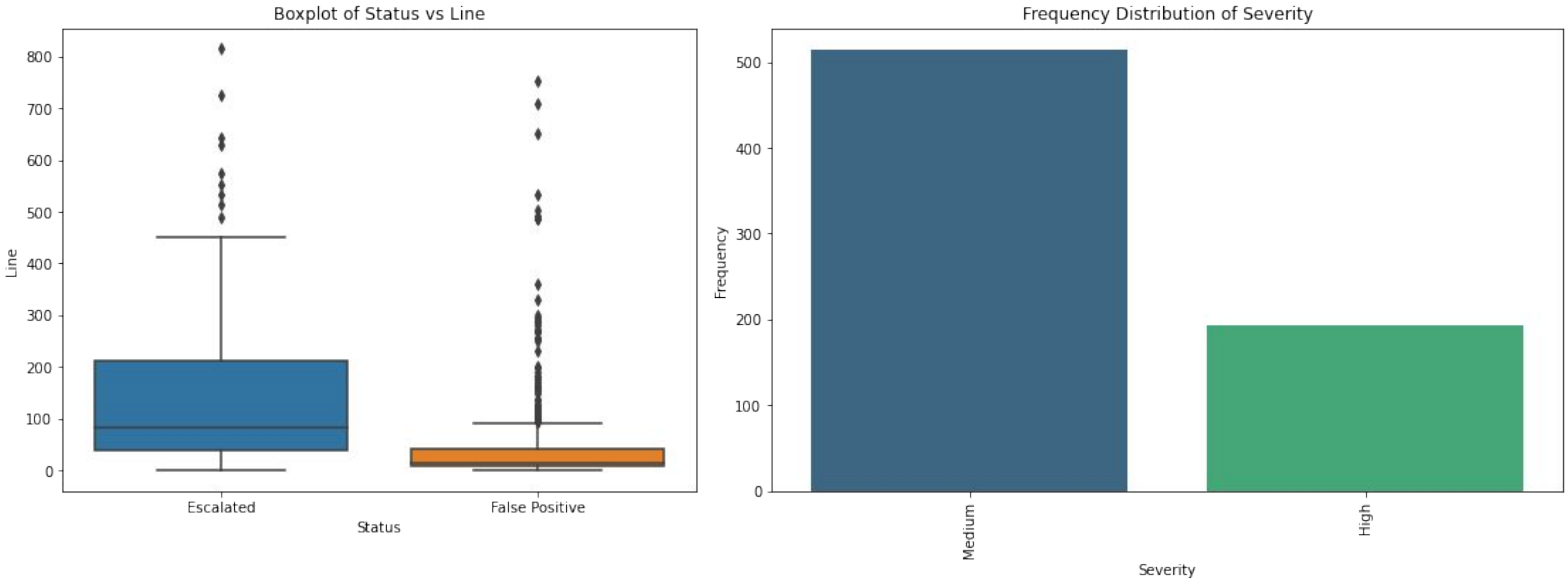
Data Cleaning: Detecting and handling missing values, outliers, and inconsistencies to ensure data quality.

Statistical Modeling: Applying statistical models like linear regression, logistic regression, decision trees, etc., to understand relationships and make predictions based on the dataset.

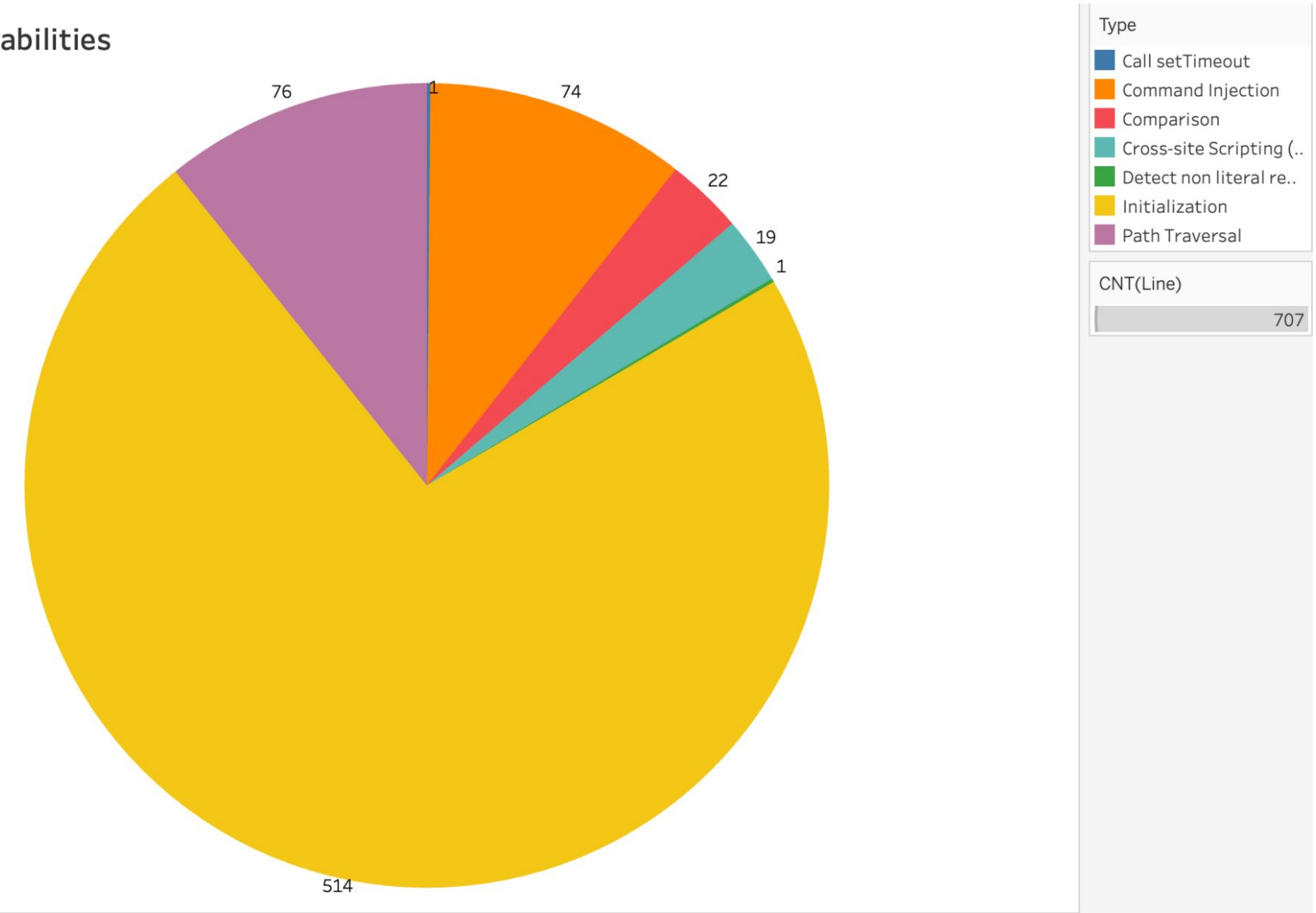
Feature Engineering: Creating new features derived from existing ones to improve model performance or gain better insights.

Machine Learning Models: Using supervised and unsupervised learning algorithms to derive patterns and make predictions or classifications based on the data.

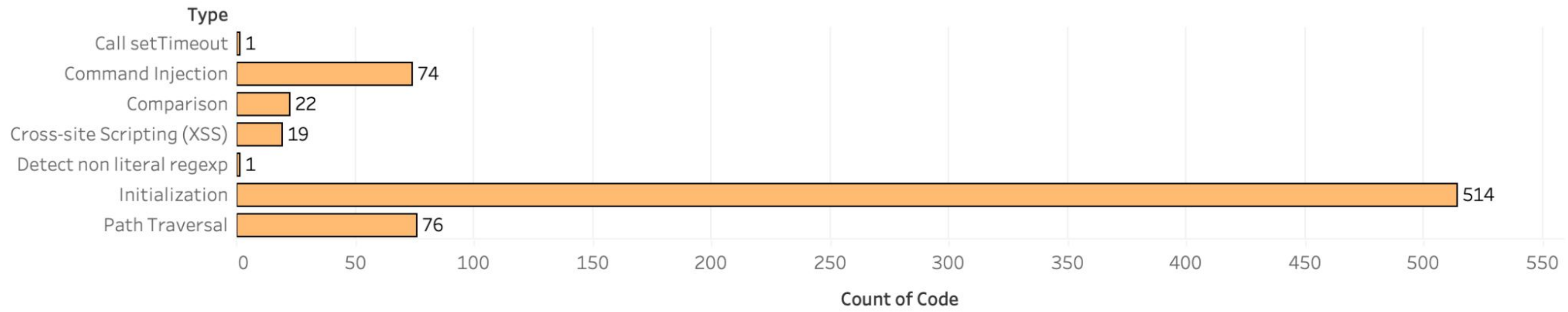
DATA ANALYSIS – EXPLORATORY DATA ANALYSIS



Types of Vulnerabilities



Count of Location by Type



Future Research

To Evaluate Threat Intelligence - We will leverage NLP to analyze unstructured data such as reports, forums, and news articles, extracting valuable threat intelligence

For Pattern Recognition - Decision trees can be used for both classification and regression tasks and are quite interpretable.

Random Forests are ensembles of decision trees that can handle larger datasets and reduce overfitting.

(K-NN)Assigns a class label to an input sample based on the majority vote of its k-nearest neighbors in the feature space.

.

CONCLUSION

- AI and ML are powerful tools that can help to improve the safety of software applications by identifying and assessing vulnerabilities more accurately and efficiently.
- We are committed to using AI and ML to improve our software and protect our users from security threats.
- We believe that AI and ML will play a vital role in the future of cybersecurity, and we are working to stay ahead of the curve.
- We are excited about the potential of AI and ML to help us create a more secure and safe digital environment for everyone. We believe that by using these technologies to identify and assess software vulnerabilities, we can reduce the risk of cyberattacks and protect our users from harm.

REFERENCES

1. SentinelOne. (2023, March 15). Advancing security: The age of AI & Machine Learning in Cybersecurity. https://www.sentinelone.com/blog/advancing-security-the-age-of-ai-machine-learning-in-cybersecurity/
2. Simplilearn. (2023, June 6). Data modeling: Overview, concepts, and types: Simplilearn. Simplilearn.com. https://www.simplilearn.com/what-is-data-modeling-article
3. Villegas, F. (2023, October 18). Data manipulation: What it is, Techniques & Examples. QuestionPro. https://www.questionpro.com/blog/data-manipulation/