Cyber Threat Intelligence Using Machine Learning

Maurice Bugg | Computer Engineer
College Park Scholars – Science & Global Change Program
mbugg215@terpmail.umd.edu | Academic Showcase, May 1, 2020

INTRODUCTION

- Every type of cyberspace is exposed to some type of cyberattack.
- On the dark web hackers post data on their most recent attacks to share with the community.
- This research project focuses on how we can use machine learning to detect cyber threats from dark web forum posts.

SERVICE SITE

Bowie State University:
https://bowiestate.edu/

1400 Jericho Rd, Bowie, MD 20715

To empower a diverse population of students to reach their potential by providing innovative academic programs.

RESULTS

- The Support Vector Machine algorithm had the highest overall accuracy rate of 85.0%.
- The Artificial Neural Network had a weighted average of 66.4%.
- The Artificial Neural Network did not report an average MCC. To be discussed later in discussion section.

	Support Vector Machine (SVM)		Artificial Neural Network (ANN)	
Exploit Type	**Accuracy**	**MCC**	**Accuracy**	**MCC**
System	95.6 %	0.698	99.9%	0.69
Web	72.8%	0.774	0.00%	
Network	58.4%	0.633	0.00%	
Database	77.3%	0.832	0.00%	
Website	70.6%	0.737	0.023%	0.13
Mobile	60.0%	0.158	0.00%	
Weighted Average	85.0%	0.695	66.4%	

DISCUSSION

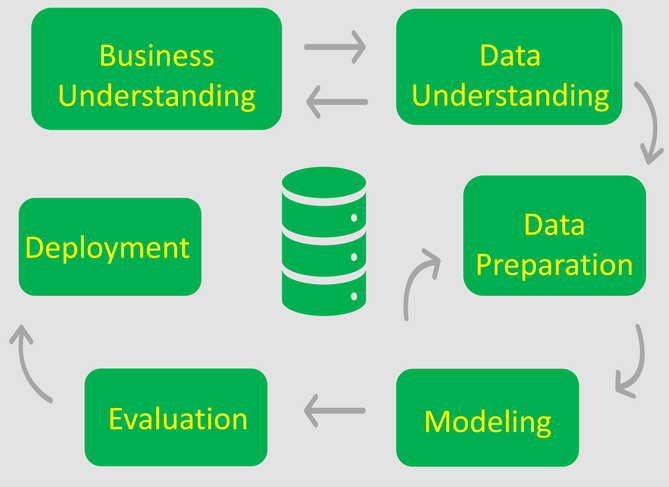
- The effectiveness of each machine learning was determined by the Average Accuracy Rate and Matthew Correlation Coefficient
- Therefore, we fail to reject the null hypothesis that the SVM models have high accuracy and reliability.
- The ANN did not report any significant percentages because it needs a lot of training data before it can make accurate models.
- Only the System type exploits reported a high accuracy because out of our 15,000+ data points 58% of them were system. This was enough for the ANN to work but the other exploits did not have enough to model.

MATERIALS

Most of our materials include bibliographic sources and archived projects which are listed in the bibliography section.

- mySQL: database management system
- Excel: clean and prepare data
- Tableau: data visualization tool
- WEKA: machine learning software to analyze data

METHODS



The infographic above shows the Cross Industry Process for Data Mining method (CRISP--DM). This is the basic methodology we followed while carrying out the research.

1. Obtained dataset of darknet forum post collected from University of Arizona
2. Filtered dataset removing duplicate instances and converted data into usable format for Tableau.
3. Uploaded data to Tableau software to visualize the data. (Not analysis yet)
4. Analyzed data using Artificial Neural Network (ANN) and Support Vector SVM machine learning algorithms from WEKA

DISCUSSION CONT.

- To transition to a proactive approach, machine learning might be the answer with its various algorithms and predictive capabilities
- Machine learning can quickly examine content posted on cyberspace and determine whether it is a threat or not

ACKNOWLEDGMENTS

I would like to acknowledge my supervisors and mentor for supporting and helping me throughout my time at Bowie State University, as well as DRs. Holtz and Merk. Also, would like to give a special thanks to:

- Mr. Pius Odhiambo for watching over us everyday, showing us what hard work and dedication really means, and providing insightful information contributing to our research.
- Ms. Angel clay for making special accommodations throughout the summer to suit all of our needs.
- Asmamaw Mersha, Christopher Clay, Overton Wright, Robel Baraki, Cion Sandidge, and Jason who were my other fellow researchers that made this experience fun and were threre throughout the entire program.

Bibliography

1. A Beginner's Guide to Neural Networks and Deep Learning. (n.d.). Retrieved July 15, 2019, from https://skymind.ai/wiki/neural-network
2. Aggarwal, M. (2018, January 07). Cross-Industry process for data mining - mayank aggarwal. Retrieved from https://medium.com/@thecodingcookie/cross-industry-process-for-data-mining-286c407132d0
3. Anticipating Cyber Vulnerability Exploits Using Machine ... (n.d.). Retrieved from https://go.recordedfuture.com/hubfs/reports/anticipating-cyber-exploits.pdf
4. Benjamin, V., Li, W., Holt, T., & Chen, H. (2015). Exploring threats and vulnerabilities in hacker web: Forums, IRC and carding shops. *2015 IEEE International Conference on Intelligence and Security Informatics (ISI)*. doi:10.1109/isi.2015.7165944
5. Classification: True vs. False and Positive vs. Negative | Machine Learning Crash Course | Google Developers. (2019, March 5). Retrieved from https://www.digitaltrends.com/cool-tech/what-is-an-artificial-neural-network/