UNIVERSITY OF ZAGREB
FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

BACHELOR THESIS num. 000

Application of NLP to threat classification from cybersecurity records

Mirta Medak

Zagreb, May 2022.

Umjesto ove stranice umetnite izvornik Vašeg rada.
Kako biste uklonili ovu stranicu, obrišite naredbu `\izvornik`.

CONTENTS

1. Introduction

Every entity that is dependent on a computer system, from corporations to individuals, could be a subject to cyber attacks.
In 2018 there were 80,000 cyber attacks per day or over 30 million attacks per year. [1] During the CoVid-19 pandemic, the cybercrime went up 600, as PurpleSec suggests. [2]
As technology evolves, more various threats to its security emerge. Tracking, describing, and evaluating these threats is of use when developing defense systems and making business decisions.

1.1. Common Vulnerabilities and Exposures System

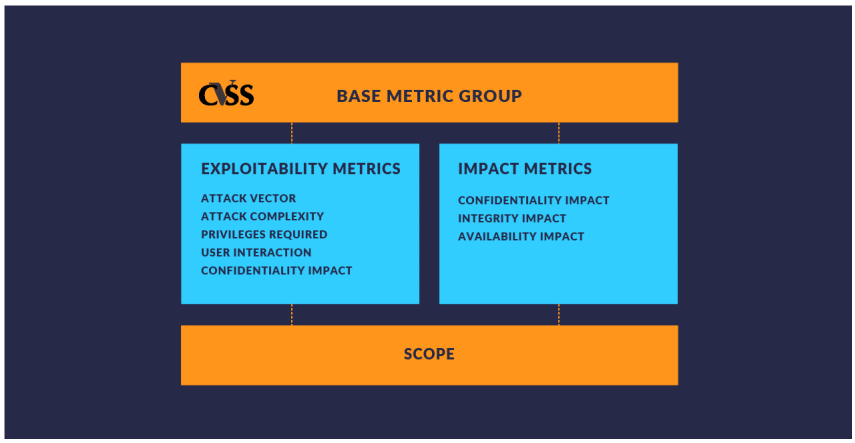
A **vulnerability** is a weakness in a computer system, that can be exploited by an attacker to execute malicious commands, access data in an unauthorized way, or perform other types of cyber attacks. [4,5]
A **threat** is any circumstance or event which has the potential to compromise system security. [6]
In order to tackle the cybersecurity problems in a more organized manner, a system of *CVE (Common Vulnerability and Exposure)* has been developed.
The *MITRE Corporation* maintains a public database of an increasing number of CVE records.
A CVE record includes an ID, a brief description of the vulnerability, and references.

1.2. Common Vulnerability Scoring System

In order to manipulate and prioritize vulnerabilities in a system, the metric of the Common Vulnerability Scoring System (CVSS) score is used. This metric estimates how "dangerous" an exploitation of a vulnerability is. CVSS is an emerging standard of vulnerability comparison. [8]
CVSS is divided in three groups: *Base, Temporal and Environmental score*. **Base Score**

group shows the traits of the vulnerability that do not change over time and are not dependent on the environment. [9] Predicting the base score will be the subject of this research.

Figure 1.1: There are eight CVSS Base submetrics:



Each submetric is assigned by experts. E. g., attack complexity can be assigned as High or Low. From these submetrics a hard-coded formula is used to compute the final Base score. Atefeh Khazaei et al. show an important concern in their work [8]: the CVSS calculation **can be subjective**. Moreover, the annotation requires experts and time, which is prolonging the process and is costly.

In this research, NLP methods are used to predict the CVSS Base score by analyzing the description from the CVE record.

2. Related work

3. Data

4. Models

5. Experiments and results

6. Conclusion

Conclusion.

BIBLIOGRAPHY

LIST OF FIGURES

Application of NLP to threat classification from cybersecurity records

Abstract

Abstract.

Keywords: Keywords.

Naslov

Sažetak

Sažetak na hrvatskom jeziku.

Ključne riječi: Ključne riječi, odvojene zarezima.