**Phase 1 - Project introduction**

**Submitted by:**

* Einav Pincu
* Idan Buller

CVE, short for Common Vulnerabilities and Exposures, is a list of publicly disclosed computer security flaws. This project analyzes the different fluctuations of the CVE in the Cyber Security world during the last decade.

For this project, we used [cvedetails.com](https://www.cvedetails.com/), which provides a web interface for CVE vulnerability data. In this website we could browse for vendors, products and versions and view CVE entries, and vulnerabilities related to them. Also, we could view statistics about vendors, products and versions of products.

CVE vulnerability data are taken from National Vulnerability Database (NVD) XML feeds provided by the National Institute of Standards and Technology. Additional data from several sources like exploits from [www.exploit-db.com](http://www.exploit-db.com/), vendor statements and additional vendor-supplied data, and Metasploit modules are also published in addition to NVD CVE data.

**Research question** - Our research question what is the connection between the severity score of security vulnerabilities and the rising of new technologies during the last decade.

Over the last decade, the technology has evolved significantly, the cyber world has also evolved and we hear about cyber attacks taking place in the political, business and individual market.

As a result, every year different security vulnerabilities are being discovered at different levels of severity and importance.

We intend to find out if the numerical value of the severity score of new security vulnerabilities may be predicted based on previous vulnerabilities data through the last decade.

**Features** – 13 columns are mentioned below:

* CVE ID (CVE-Year-ID)
* Vulnerability Type (Types of vulnerabilities such as Bypass, DOS, XSS, etc.)
* Publish Date (Date)
* Update Date (Date)
* Score (Severity of vulnerability – 1.0-10.0)
* Gained Access Level (None \ Partial \ Complete)
* Access (Local \ Remote)
* Complexity (Low \ Medium \ High)
* Authentication (Required \ Not required)
* Confidentiality (None \ Partial \ Complete)
* Integrity (None \ Partial \ Complete)
* Availability (None \ Partial \ Complete)

**Instances** –174,954 rows, each row contains information about a vulnerability.

**Data sources** –

[cvedetails.com](https://www.cvedetails.com/) - Security vulnerability database/information source. In that website we can view various details (Vulnerability score, access level, complexity, integrity, etc.) regarding vulnerabilities throughout the years,

**Data mining methods** – Crawling and scraping from CVE Details.

**Machine learning model** - Our Machine learning model is Linear Regression with which we will predict the CVE Score of various vulnerabilities.

**Validation methods** – Our validation model is R2.