**Data science project**

Submitted by:  
Einav Pincu – 315113175  
Idan Buller – 208713792  
**Research question** - The question is what is the connection between the severity score of security vulnerabilities and the rising of new technologies in the web surface.

Over the last decade, web technology has evolved significantly, enterprises are developing new services which expand the exposure of the enterprises' services to the internet.

At the same time the cyber and information security field has adapted itself. As a result, every year different vulnerabilities are being discovered at different levels of severity and importance.

**We intend** to find out if the numerical value of 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** – over 50,000 rows, each row will contain information about a vulnerability.

**Data sources** –

CVE Details - 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.  
<https://www.cvedetails.com/>

VulnDB - Comprehensive and timely vulnerability intelligence which provides actionable information about the latest vulnerabilities in security field.  
<https://vulndb.cyberriskanalytics.com/>

Rapid7 - Technical details for over 180,000 vulnerabilities.   
<https://www.rapid7.com/db/>

**Data mining methods** – API and crawling from various websites such as: CVE Details, VulnDB, Rapid7. Libraries – BeautifulSoup, Requests, pandas.

**Planned visualizations** – 2D visualizations, based on different methods which we have learned during our course. For example: pie charts, scatter plots etc.

**Planned models** - Linear Regression.

**Validation methods** – R2.