

Plague inc.

Burián Sándor

Molnár Vince



ÓBUDAI EGYETEM
ÓBUDA UNIVERSITY

Tartalom

Bevezetés.....	3
Adatok	3
Használt módszertan	4
Felhasznált eszközök	5
Microsoft Azure	5
Adatok kezelése.....	6
Adatvizualizáció és UI	7
Összefoglalás	9
Bibliográfia.....	10
Képjegyzék.....	10
Képletjegyzék.....	10

Bevezetés

Március végén részt vettünk a Hack&Heal című 48 órás hackathonon¹ a crafthub szervezésében, ahol számos remek egészségügyi ötlettel találkoztunk és számos nagy tudású mentorral tudtunk beszélni. A mi projektünk, matematikus és adatszaki mentorokkal egyeztetve egy járvány előrejelző rendszer alapjainak megteremtése volt, azaz a szükséges feltételek megteremtése ahhoz hogy egy új esemény bekövetkeztekor előre megjósolhassuk, hogy mekkora mértékben fog hatással lenni egyes országokra/területekre. Természetesen egy ekkora feladatot ennyi idő alatt nem lehet elvégezni, ezért ennek csak a létszükségletének feltételeit kipróbálandó és működésének létjogosultságát bizonyítandó egy kis próbaverziót készítettünk el. Ehhez nagy mértékű adatkeresést folytattunk és a lehetséges matematikai modelleket kikerestük, majd vizualizációs technikákkal foglalkoztunk felhőre optimalizálva, hogy a hatékonyságát növelhessük. Mivel az előrejelzéshez mi csak bebizonyítani szerettük volna, hogy az ötlet működik, ezért korábbi járványokat vettünk alapul és vetettük össze korábbi természeti katasztrófákkal.

A feladatot modellezendő Kamenár Dáviddal készítettünk egy rövid YouTubeon is elérhető videót² egy lehetséges use-caseről.

Adatok

Leő és legfontosabb kérdés az elegendő mennyiségű és minőségű adat volt. Ezért, hogy leszűkítsük a kört az elmúlt évszázad legfertőzőbb és legnagyobb halálozási rátával jellemezhető járványait vettük alapul [1]. Ugyanakkor fontos megemlíteni, hogy a lehetséges veszélyes vírusokként számontartott WHO lista például a CDC időről-időre kiadott jelentései alapján nem feltétlen vehető az egyetlen kiindulási alapnak, hiszen a CDC-nél már 2017-ben megjelent a covid mint lehetséges veszélyforrás [2].

A fegyveres atrocitásokra ugyan találtunk adatokat, de mivel a második világháború óta csupán 26 nap telt el békében[3] a Földön így nem találtuk kielégítőnek ezért nem szerepeltetjük a demóban.

Ezek után leszűkítettük a modellezendő betegséget a lassa-láz – zika - ebola bármásra, de mivel a lassa-lázhoz jelenleg még nagyon kevés adat áll rendelkezésre, és az ebola

¹ <https://hacknheal.crafthub.events> (2022)

² Plague Inc - a lassa fever story; <https://youtu.be/KxBMI9pYkIE> (2022)

feljegyzéseinek jelentős része offline francia nyelvű a gyarmati időkben így jelen esetben csak a Zika esetein szeretnénk bemutatni a módszert, ugyanakkor a meglévő adatainkon természetesen működik a többi is, de kiemelnénk, hogy nem tartjuk jelen állapotában ielégítőnek a kapható eredményeket.

Ezért a használt adatok a CDCTől származnak a Zika vírusra[4], az EM-DAT adatbázisát használtuk természeti katasztrófákhoz [5] és szintén a CDC[6] és a félhivatalos wikipédán is elérhető[7] adatbázist használtuk Ebolára.

Használt módszertan

Az alap ötletünk az volt hogy Big data technológiák használatával próbáljuk előre jelezni a lehetséges járványok előfordulását. Annyi különböző, de valamilyen szinten összefüggő adatforrás használatával amennyivel lehetséges. De úgy gondoljuk, példaként használhatjuk a középkori Európa adatait néhány modern fejlődő ország esetében is. Továbbá úgy gondoljuk, hogy nem az idő, hanem a kontextus határozza meg, hogy mi használható és mi nem épp ezért az időt mint faktort csak a vizualizáció szempontjából tartottuk fontosnak. Olyan metrikát próbáltunk keresni ami a mi esetünkben jól használható.

Erre legmegfelelőbb modellként[8] az incidencia számítás[9] tartottuk, mivel ebben a esetben tudjuk az adott időintervallumra (természeti katasztrófa/háború ideje, vagy az azt követő időszak; *a vizualizáció során nem számoltunk rákövetkező időszakokkal, ugyanakko egy jelentős és fontos továbbfejlesztési lehetőség!*) leszűkíteni azt, hogy az ekkor számolt új betegeknek van-e kapcsolata az ekkor bekövetkezett eseményekkel.

$$\frac{\text{új esetek száma vizsgált periódusban}}{\sum_1^{\text{összes megfigyelt személy}} \text{megfigyelt személy}}$$

egyenlet 1 incidencia számítására használt képlet

Kiemelnénk, hogy természetesen ez a modell csak akkor működik megfelelően, ha minél több faktort tudunk benne egyesíteni, ugyanakkor a hackathon időkorláta miatt mi számos jelentős faktort nem vettünk figyelembe, így természetesen a vizualizáció során fenn áll a confounding esete.

Ilyen faktorok lehetnek például a csapást elszenvedő terület kapcsolata azokkal ahol esetek jelentkeznek, mind távolság mind közlekedési lehetőség szempontjából. Hasonló lehet az adott

társadalmi berendezkedése vagy gazdasági állapot és nyilvánvalóan az egészségügy mint rendszer állapota is.

Ezen kívül az adott járványhoz ideális esetben az adott betegség tulajdonságaival lenne érdemes számolnunk, mint a lappangási idő, fertőzési ráta, halálozási arány, stb

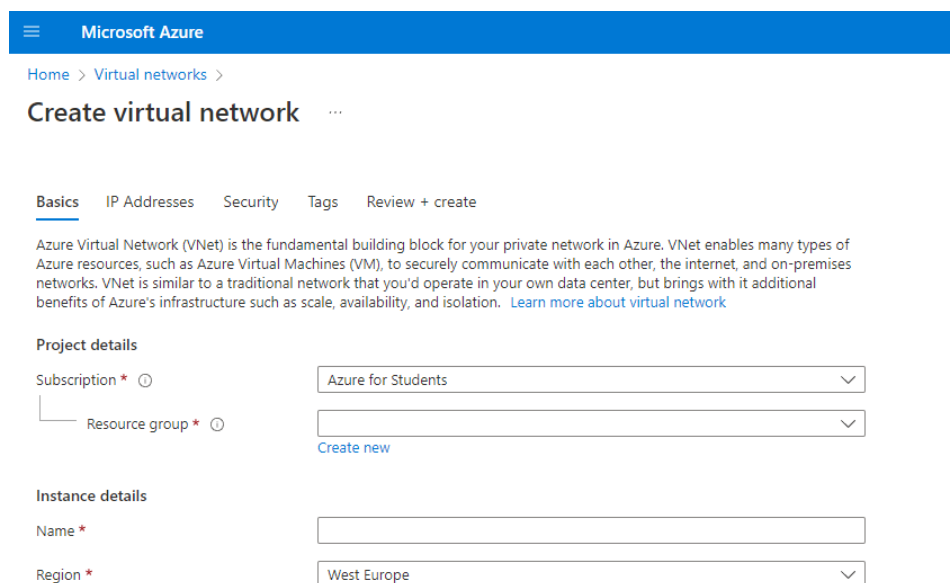
A modell továbbfejlesztésére javasoljuk továbbá az időár és évszak modellek alkalmazását is, amit a járvány tulajdonságaival összevetve képesek vagyunk kiszűrni a false positive eseteket a modell szempontjából.

Felhasznált eszközök

Microsoft Azure

Először egy Resource group létrehozására volt szükség, ami egy container Azure környezetben.

Ezután Azure Virtual Network létrehozására volt szükség. Ami a magánhálózat alapvető építőköve az Azure-ban. A virtuális hálózat számos típusú Azure-erőforrást tesz lehetővé, például az Azure Virtual Machines számára, hogy biztonságosan kommunikáljanak egymással, az internettel és a helyi hálózatokkal.

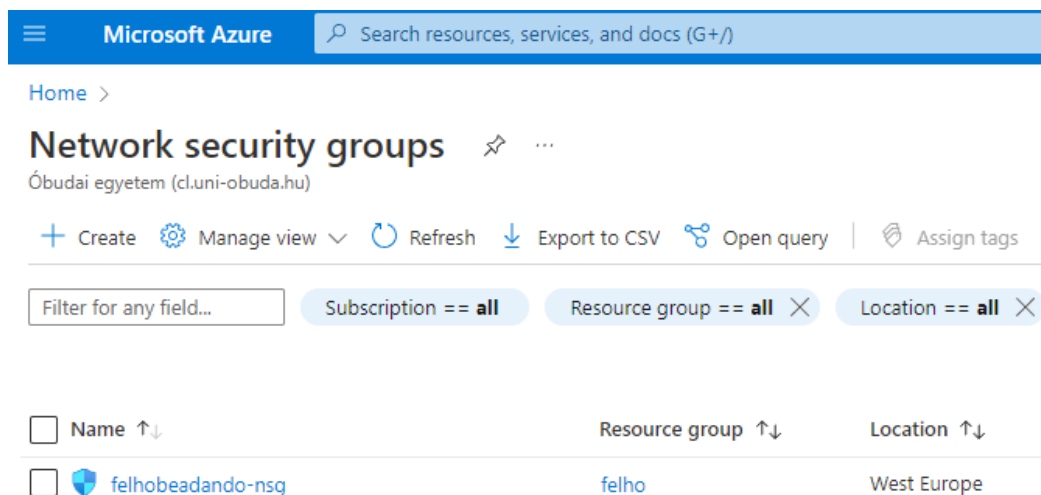


The screenshot shows the 'Create virtual network' wizard in the Microsoft Azure portal. The 'Basics' tab is active. The form contains the following fields:

- Subscription ***: A dropdown menu showing 'Azure for Students'.
- Resource group ***: A dropdown menu with a 'Create new' link below it.
- Name ***: A text input field.
- Region ***: A dropdown menu showing 'West Europe'.

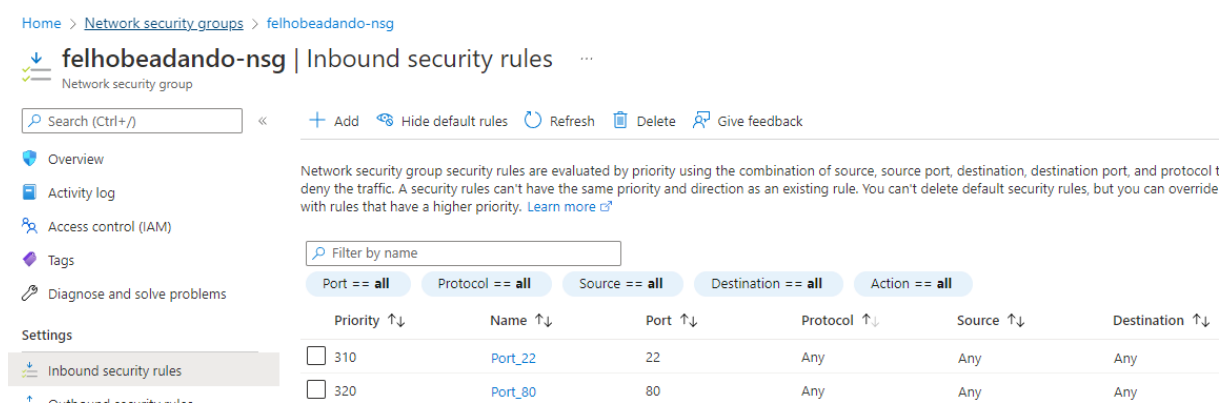
ábra 1 virtuális hálózat létrehozása

Következő lépés egy Network Security Group létrehozása. Aminek segítségével lehet szűrni az erőforrástól érkező, vagy az erőforrás felé irányuló forgalmat, amik a Virtual Network-ön keresztül haladnak. Összességében egy hozzáférés vezérlési szabályok készlete.



ábra 2 security group konfigurálás

Ehhez a Security grouphoz két új Inbound security rule-t adtunk.



ábra 3 security rule hozzáadása

22-es porton az SSH elérhetőség érdekében, 80-as porton pedig FTP elérhetőség miatt.

Ezután következett az Azure Virtual Machine létrehozása.

Itt kiválasztottuk a már létrehozott Resource group-ot és Virtual network-öt, valamint kiválasztottuk a kívánt Image-et, ami a mi esetünkben egy Ubuntu server 20.04 LTS-Gen2.

A VM létrehozása után SSH-n keresztül rácsatlakoztunk, és telepítettük az Apache servert.

Adatok kezelése

Az adatokat egy-egy JSON állományként kezeltük, amiket lokálisan töltünk be a szerverről és kezeli le a front-end oldala. Ez külön-külön datasetet jelent a járványok és természeti katasztrófák – háborúk esetén a könnyebb utólagos módosítás kezelés végett illetve a különböző adatforrások okán is.

Az adatokat természetesen nem ilyen formátumban találtuk meg, így át kellett alakítani minden esetben ami egy-egy nagy adathalmaz esetén jelentős mennyiségű művelet, így azt inkább felhőben végeztük el pár segédscript használatával.

Adatvizualizáció és UI

Az adatvizualizációhoz a Google Charts – Geo chart³ rendszert használtuk, amivel szép és egyszerű de a fleadatnak megfelelő eredményeket lehet elérni.

A UI felépítéséhez szintén javascriptet és CSS-t használtunk Zurb-Foundation⁴ keretrendszerrel, mivel kevés adatot tölt be és kényelmes mobil barát felületet hoz létre.

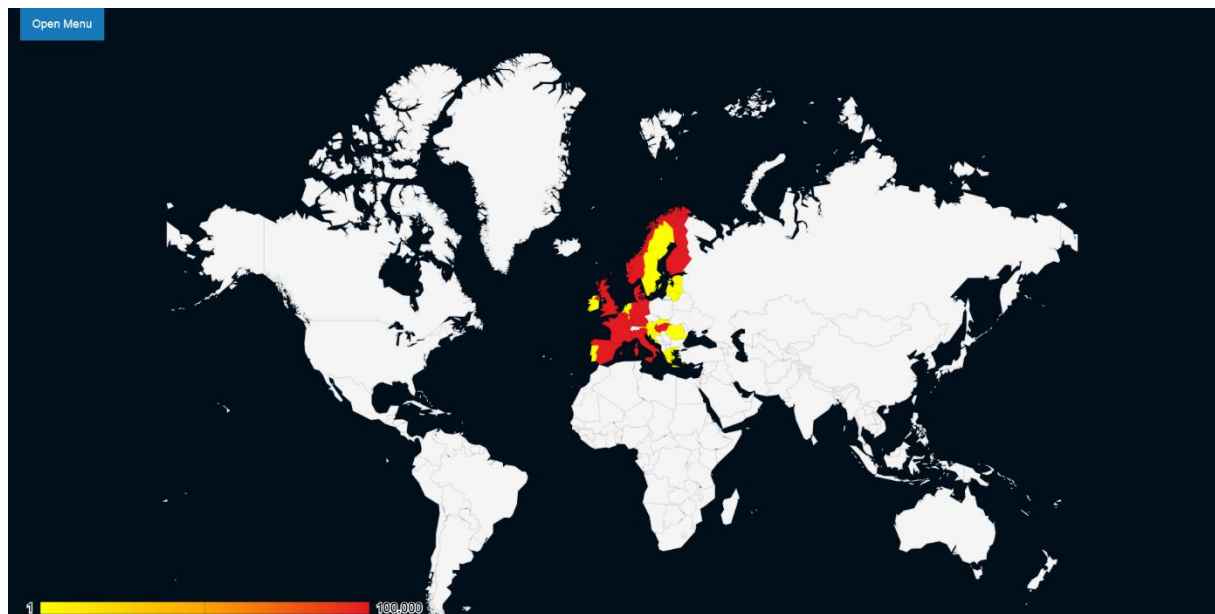
Így az alábbi eredményeket tudtuk elérni:



ábra 4 2016 legutolsó természeti csapásának hatása az ismert Zika esetekkel összevetve ugyanazon időszakban

³ <https://developers.google.com/chart/interactive/docs/gallery/geochart> (2022)

⁴ <https://get.foundation/> (2022)



ábra 5 2019 legutolsó természeti csapásának hatása az ismert Zika esetekkel összevetve ugyanazon időszakban



ábra 6 a vezérlőpanel

Összefoglalás

Összefoglalva megmutattuk, hogy a modell működik és képesek vagyunk használni ilyen természetű adatokat járványok modellezésére.

A várható járványok előrejelzésére a fenti módszerben a javasolt modell kiegészítéseket követően akár valós időben is tudunk modellezni eseményeket, így például egy vulkán kitörés pillanatában megmondható, hogy várhatóan hol és mit fog okozni olyan szintekig a társadalmi struktúrában, amit jelenleg elképzelni sem tudunk.

Ugyanazén módszert történelmi adatokra vetítve viszont akár a múlt egyes eddig ismeretlen eseményei is megvilágosodnak.

Bibliográfia

- [1] „List of epidemics”, *Wikipedia*. 2022. május 2. Elérés: 2022. május 3. [Online]. Elérhető: https://en.wikipedia.org/w/index.php?title=List_of_epidemics&oldid=1085860732
- [2] „NNDSS - Table I. infrequently reported notifiable diseases | Data | Centers for Disease Control and Prevention”, *Data.CDC.gov*. <https://data.cdc.gov/NNDSS/NNDSS-Table-I-infrequently-reported-notifiable-dis/45b4-9j7u> (elérés 2022. május 3.).
- [3] Hír TV, *Célpont (2021-05-01) - HÍR TV*, (2022. május 1.). Elérés: 2022. május 3. [Online Video]. Elérhető: <https://www.youtube.com/watch?v=5KTjJbcNwzY>
- [4] CDC, „Reporting and Surveillance - Zika Virus”, *Centers for Disease Control and Prevention*, 2017. január 12. <http://www.cdc.gov/zika/reporting/index.html> (elérés 2022. április 24.).
- [5] „EM-DAT | The international disasters database”. <https://emdat.be/> (elérés 2022. április 24.).
- [6] „History of Ebola Virus Disease (EVD) Outbreaks | History | Ebola (Ebola Virus Disease) | CDC”, 2022. február 18. <https://www.cdc.gov/vhf/ebola/history/chronology.html> (elérés 2022. április 24.).
- [7] „List of Ebola outbreaks”, *Wikipedia*. 2022. január 30. Elérés: 2022. április 24. [Online]. Elérhető: https://en.wikipedia.org/w/index.php?title=List_of_Ebola_outbreaks&oldid=1068907322
- [8] CDC, „Principles of Epidemiology | Lesson 3 - Section 2”, 2012. május 18. <https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section2.html> (elérés 2022. április 24.).
- [9] Will Kenton, „What Does the Incidence Rate Measure?”, *Investopedia*, 2021. szeptember 25. <https://www.investopedia.com/terms/i/incidence-rate.asp> (elérés 2022. április 24.).

Képjegyzék

ábra 1 virtuális hálózat létrehozása.....	5
ábra 2 security group konfigurálás	6
ábra 3 security rule hozzáadása.....	6
ábra 4 2016 legutolsó természeti csapásának hatása az ismert Zika esetekkel összevetve ugyanazon időszakban.....	7
ábra 5 2019 legutolsó természeti csapásának hatása az ismert Zika esetekkel összevetve ugyanazon időszakban.....	8
ábra 6 a vezérlőpanel	8

Képletjegyzék

egyenlet 1 incidencia számítására használt képlet	4
---	---