

**SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE  
FAKULTA ELEKTROTECHNIKY A INFORMATIKY**

**PREDIKCIA EPIDEMIOLOGICKÝCH DÁT  
V SÚVISLOSTI S PANDÉMIOU COVID-19 METÓDAMI  
STROJOVÉHO UČENIA A NEURÓNOVÝMI SIEŤAMI**

**TÍMOVÝ PROJEKT**

**Bc. Filip Frank  
Bc. Marek Kačmár  
Bc. Viet Quoc Le  
Bc. Juraj Puzster  
Bc. Tomáš Singhofer  
Bc. Branislav Šipula  
Bc. Adam Vozár**

**Zadávatel': MUDr. Veronika Kurilová, PhD.**

# Obsah

<b>Úvod .....</b>	<b>1</b>
<b>1 Zbierané dáta na Slovensku v súvislosti COVID-19 .....</b>	<b>2</b>
<b>2 Microsoft Azure .....</b>	<b>3</b>
2.1 Azure Machine Learning .....	3
2.2 Príprava dát pre tréningovanie modelov pomocou AutoML .....	5
<b>3 Modely strojového učenia tréňované pomocou AutoML .....</b>	<b>6</b>
3.1 Seasonal Naive.....	6
3.2 ARIMA a ARIMAX .....	6
3.3 Decision Tree.....	7
3.4 Random Forest.....	8
3.5 XGBoost .....	9
3.6 Voting Ensemble .....	10
3.7 Stacking Ensemble.....	11
3.8 Elastic Net.....	11
<b>4 Doterajšie výsledky.....</b>	<b>13</b>
4.1 Predikcia pozitívnych PCR testov .....	13
4.2 Predikcia pozitívnych Ag testov .....	14
4.3 Predikcia úmrtí z viacerých ukazovateľov.....	15
4.4 Predikcia úmrtí z úmrtí .....	16
4.5 Predikcia kompletne zaočkovaných.....	17
4.6 Graf celkových hospitalizácií .....	18
4.7 Graf celkových UPV.....	19
<b>Zoznam použitej literatúry .....</b>	<b>20</b>

# Zoznam obrázkov

Obrázok 1 Ukážka csv formátu datasetu IZA.....	2
Obrázok 2 Životný cyklus projektu .....	4
Obrázok 3 Znáznornenie služby Azure Machine Learning .....	4
Obrázok 4 Rovnica Seasonal Naive .....	6
Obrázok 5 Vzorec pre model ARIMA.....	7
Obrázok 6 Vzorec pre model ARIMAX.....	7
Obrázok 7 Vytvorenie modelov ARIMAX a SARIMAX.....	7
Obrázok 8 Ukážka modelu Decision Tree.....	8
Obrázok 9 Model Random Forest .....	9
Obrázok 10 Jednoduchá ukážka princípu gradient boosting metódy s využitím rozhodovacích stromov. ....	10
Obrázok 11 Voting Ensemble .....	11
Obrázok 12 Graf časového vývoja počtu pozitívnych PCR testov (modrým) so 14-dňovou predikciou (červeným) .....	13
Obrázok 13 Graf predikcie Ag z PCR, PCR%, Ag, Ag%.....	15
Obrázok 14 Graf predikcie úmrtí z PCR, PCR%, Ag, Ag%.....	16
Obrázok 15 Graf predikcie úmrtí z úmrtí .....	17
Obrázok 16 Graf predikcie kompletne zaočkovaných.....	17
Obrázok 17 Graf celkových hospitalizácií.....	18
Obrázok 18 Graf časového vývoja UPV .....	19

# **Zoznam skratiek a značiek**

Ag – Antigénový test

CSV – Comma Seperated Value

IZA – Inštitút Zdravotných Analýz

OTP – režim očkování, testování, překonání

PCR - polymerázová reťazová reakciu s reverznou transkripciou

UPV - umelé pľúcne ventilácie

# Úvod

*(Spracoval: Marek Kačmár, upravil: Filip Frank)*

Pandémia spôsobená koronavírusom COVID-19, ktorý bol identifikovaný v decembri 2019 v Číne, zasiahla takmer všetky aspekty našich každodenných životov. Vlády jednotlivých štátov sveta neustále zavádzajú rôzne opatrenia v snahe zmierniť dopad pandémie na ekonomiu a demografiu štátu.

Medzi spomínané opatrenia patrí napríklad prechod škôl na dištančnú formu výučby, nosenie rúšok sa stalo povinným, režim OTP a i.. Vzhľadom na negatívny dopad týchto opatrení je potrebné sledovať ich efektivitu, ktorá sa prejavuje na časovom vývoji pandemických ukazovateľov, akými sú mimo iných počet vykonaných testov, podaných vakcín či obsadenosť lôžok v nemocniciach. Metódy strojového učenia majú potenciál krátkodobo predpovedať vývoj týchto ukazovateľov, čo môže uľahčiť adekvátne rozhodovanie pri riešení pandémie.

V tomto tímovom projekte skúmame možnosť využitia cloud computing platformy Microsoft Azure na tréning a nasadenie regresných modelov pre predpoveď vývoja ukazovateľov pandémie ochorenia COVID-19 na Slovensku. Naším cieľom je zoznámiť sa s prostredím platformy Microsoft Azure a následne s využitím verejne dostupných dát Inštitútu Zdravotných Analýz natréňovať a nasadiť čo najpresnejšie predikčné modely pre predpoveď vybraných pandemických ukazovateľov.

# 1 Zbierané dáta na Slovensku v súvislosti COVID-19

(Spracovali: Branislav Vozár a Adam Šipula, upravil: Filip Frank)

Hlavným zdrojom dát pre trénovanie našich regresných modelov je Inštitút zdravotníckych analýz, ktorý je analytickým a poradným útvarom Ministerstva zdravotníctva SR. Ide o dáta dokumentujúce časový priebeh pandémie COVID-19 na Slovensku, ktoré sú dostupné na <https://github.com/Institut-Zdravotnych-Analyz/covid19-data>

Dáta sú dostupné v tabuľkovom formáte, sledujeme v nich najmä časový vývoj týchto pandemických ukazovateľov:

- denný počet pozitívnych PCR a antigénových testov
- denné percento pozitívnych PCR a antigénových testov
- denný počet úmrtí
- počet hospitalizovaných pacientov
- počet pacientov na umelej pľúcnej ventilácii

	A	B	C
1	Datum	AgPosit	AgPosit%
2	11.10.2020	0	0.0
3	12.10.2020	40	7.6
4	13.10.2020	84	7.3
5	14.10.2020	31	3.3
6	15.10.2020	36	9.7
7	16.10.2020	27	8.4
8	17.10.2020	0	0.0
9	18.10.2020	0	0.0
10	19.10.2020	69	8.7
11	20.10.2020	38	6.2
12	21.10.2020	48	5.0
13	22.10.2020	141	6.6
14	23.10.2020	119	5.5
15	24.10.2020	65	5.4
16	25.10.2020	40	5.5
17	26.10.2020	231	10.7

Obrázok 1 Ukážka csv formátu datasetu IZA

## 2 Microsoft Azure

*(Spracoval: Juraj Puszter)*

Microsoft Azure je takzvaná cloud computing platforma. Poskytuje 4 rôzne formy cloud computingu a to Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Software as a Service (SaaS) a tiež serverless.

Microsoft rozdeľuje Azure cloud služby do množstva kategórií, vrátane: Compute, Mobile, Web, Storage, Analytics, Networking, Media and content delivery network, Integration, Identity, Internet of things, DevOps, Development, Security, Artificial intelligence and machine learning, Containers, Databases, Migration, Management and governance, Mixed reality, Blockchain, Intune.

Vlastnosti Microsoft Azure:

- Flexible – ľubovoľné presúvanie výpočtových zdrojov
- Open – podporuje takmer akýkoľvek operačný systém, jazyk, nástroj alebo framework
- Reliable – 99.95% dostupnosť SLA a 24x7 technická podpora
- Global – dáta sú uložené v geosynchronných dátových centrách
- Economical – platí sa iba za to, čo sa používa

### 2.1 Azure Machine Learning

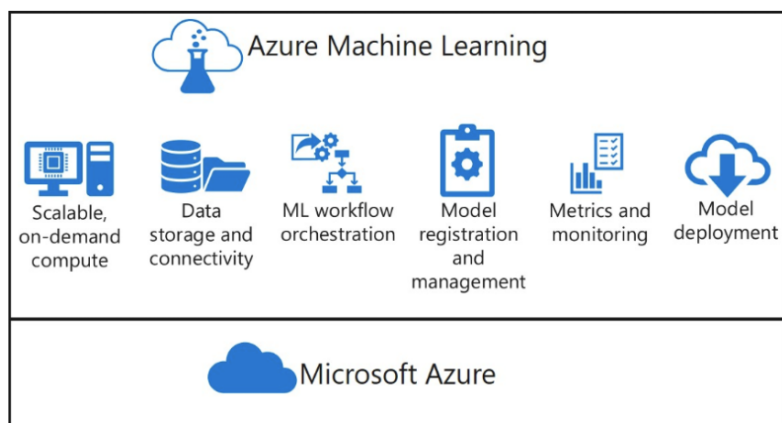
**Azure Machine Learning** je cloudová služba pre urýchlenie a správu životného cyklu projektov strojového učenia. Používa sa pre tréning, nasadzovanie modelov a správu MLOps. V Azure Machine Learning je možné vytvoriť model alebo použiť model z open source platformy, ako je napríklad Pytorch, TensorFlow alebo scikit-learn. MLOps nástroje pomáhajú s monitorovaním, pretrénovaním alebo znovu nasadením modelov.

**Azure Machine Learning Studio** je web portál pre ľudí využívajúcich Azure Machine Learning. Tento web portál uľahčuje a poskytuje prácu s Azure Machine Learning priamo cez internetový prehliadač.

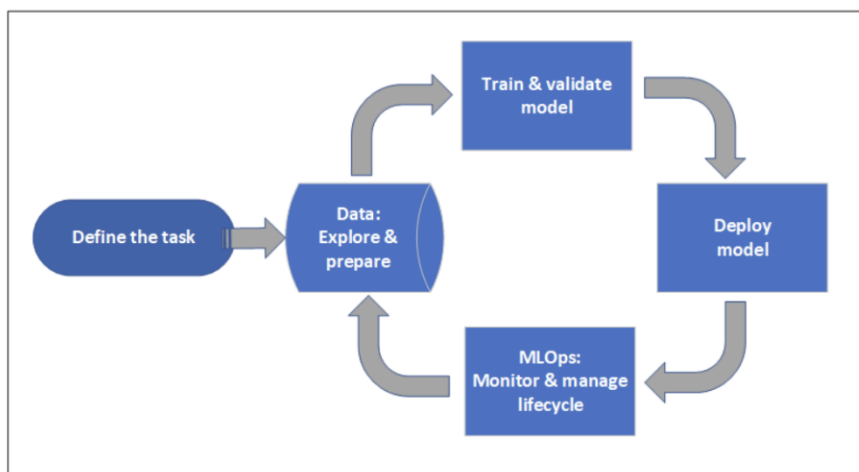
**Automated Machine Learning (AutoML)** je proces automatizácie časovo náročných, opakujúcich sa úloh vývoja modelu strojového učenia.

AutoML poskytuje 3 typy tréningu modelov. Sú to Classification, Regression a Time-Series Forecasting.

- Classification – rozdeľuje dáta do kategórií
- Regression – predvída numerickú hodnotu dát na základe nezávislých prediktorov
- Time-Series Forecasting – podobný ako regresia, ale dáta sú závislé na čase



Obrázok 3 Znáozornenie služby Azure Machine Learning



Obrázok 2 Životný cyklus projektu

Služba Azure Machine Learning podporuje okrem automatizovaného učenia pomocou AutoML aj **code-first** riešenia pre tréning modelov, ako aj návrh nasadenie modelov pomocou **drag-n-drop Azure Machine Learning dizajnéra**.



## 2.2 Príprava dát pre tréovanie modelov pomocou AutoML

(Spracoval: Filip Frank)

Vstupné dáta pre tréovania získavame z repozitára IZA pomocou vlastných jednoduchých skriptov písaných v jazyku Python. Extrahované dáta majú formu CSV súborov a v porovnaní s pôvodnými dátami prešli týmito úpravami:

1. odstránenie nerelevantných stĺpcov
2. odstránenie riadkov obsahujúcich NA hodnoty
3. aplikácia 7-dňového moving average na stĺpce tabuľky (okrem nezávislej premennej dátum).

Príklady vstupných datasetov (každý uvedený dataset obsahuje aj nezávislú premennú dátum):

1. Ag, %Ag
2. PCR, %PCR
3. Ag, %Ag, PCR, %PCR
4. Ag, %Ag, PCR, %PCR, Úmrtia
5. Úmrtia

## 3 Modely strojového učenia tréňované pomocou AutoML

### 3.1 Seasonal Naive

(Spracoval: Marek Kačmár)

Seasonal naive je jeden z mnoho modelov, ktoré používa Microsoft Azure pri predpovedi daného pripraveného datasetu. Tento model je veľmi jednoduchý, no napriek tomu efektívny.

Používa sa hlavne pri sezónnych dátach, t. j. každá predpoveď je rovná poslednej pozorovanej hodnote v rovnakej „sezóne“ v danom roku (napr. pri predpovedaní hodnôt z mesiaca Február si model bude brať hodnoty z Februára minulého roku). Jeho rovnica vyzerá nasledovne:

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)},$$

Obrázok 4 Rovnica Seasonal Naive

kde  $m$  predstavuje hodnotu sezónnej periódy a  $k$  je výsledné číslo výpočtu  $(h-1)/m$  – počtu kompletných rokov v predpovedi pred stanoveným časom  $T + h$ . Model má najlepšiu efektivitu, keď je použitý pri ekonomických alebo finančných predpovediach.

### 3.2 ARIMA a ARIMAX

(Spracoval: Juraj Puszter)

ARIMA je skratka pre Auto Regressive Integrated Moving Average. Je to trieda modelov ktorá predpovedá zadané časové rady (time series) na základe svojich vlastných minulých hodnôt. Auto regressive v názve znamená, že je to lineárny regresný model.

Model je charakterizovaný 3 výrazmi  $p$ ,  $d$ ,  $q$ .

- $p$  - Odkazuje na počet oneskorení (lags)  $Y$  ktoré sa majú použiť ako predikáty.
- $q$  - Odkazuje na počet oneskorených chýb predpovede (lagged forecast errors), ktoré by mali ísť do ARIMA modelu.

- d – Minimálny počet rozdielov potrebných na to, aby séria bola stacionárna.

$$\Delta^D y_t = \sum_{i=1}^p \phi_i \Delta^D y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t$$

$$\epsilon_t \sim N(0, \sigma^2)$$

Obrázok 5 Vzorec pre model ARIMA

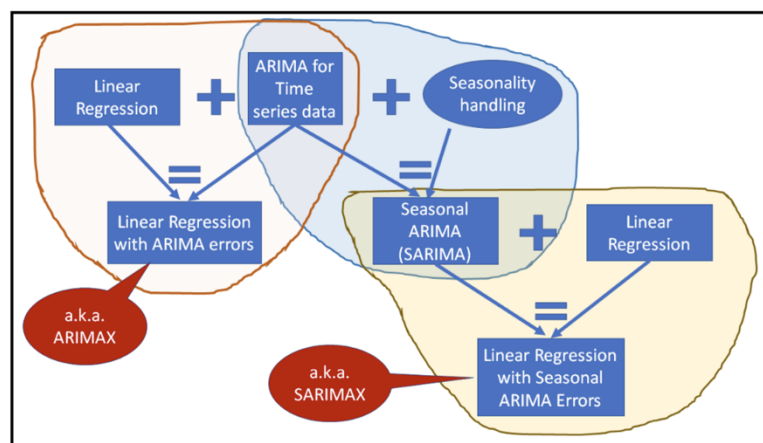
Model ARIMAX je rozšírený model ARIMA. Obsahuje ďalšie nezávislé premenné (prediktory). Model sa tiež uvádza ako vektor ARIMA alebo dynamický regresný model.

$$\Delta^D y_t = \sum_{i=1}^p \phi_i \Delta^D y_{t-i} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \sum_{m=1}^M \beta_m X_{m,t} + \epsilon_t$$

$$\epsilon_t \sim N(0, \sigma^2)$$

Obrázok 6 Vzorec pre model ARIMAX

Nasledujúci obrázok znázorňuje, ako je kombinovaná lineárna regresia, ARIMA a SARIMA pre vytvorenie modelov ARIMAX a SARIMAX.

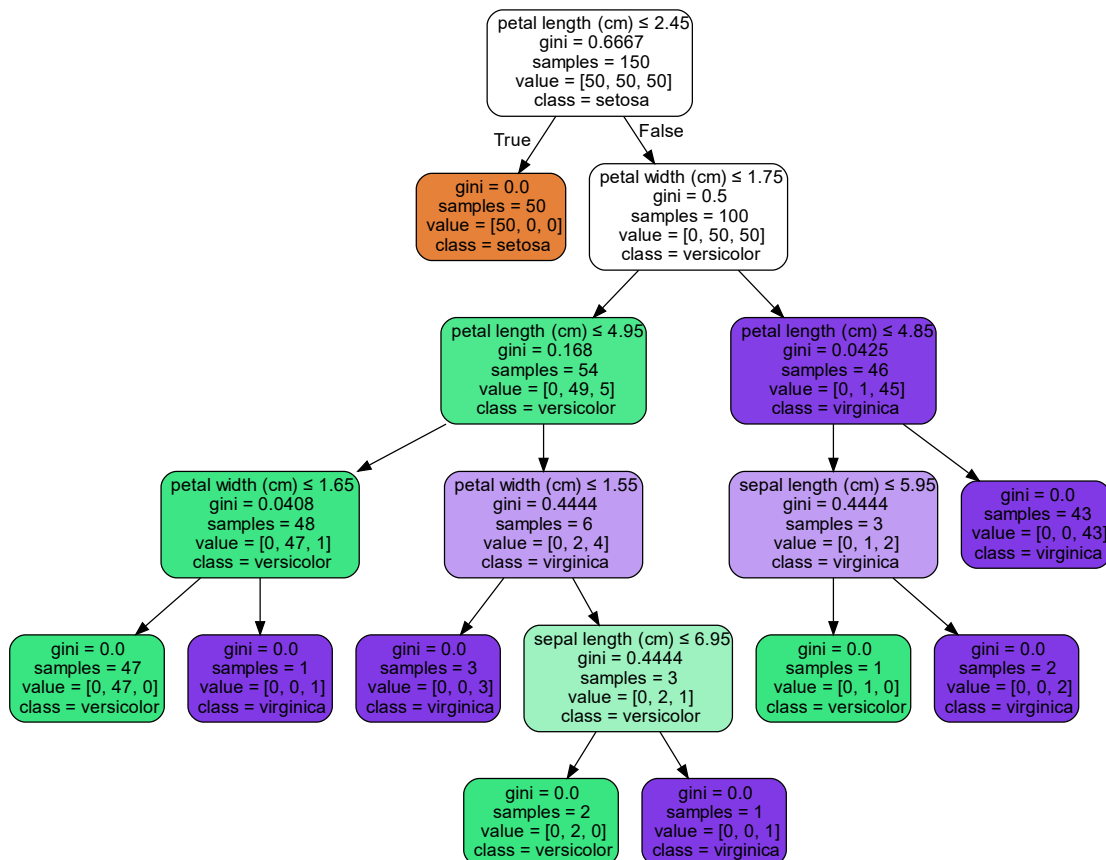


Obrázok 7 Vytvorenie modelov ARIMAX a SARIMAX

### 3.3 Decision Tree

(Spracoval: Adam Vozár)

- Učenie s učiteľom
- Jednoduchá implementácia
- If-else pravidlá
- Regresia aj klasifikácia
- Nie je potrebná normalizácia dát

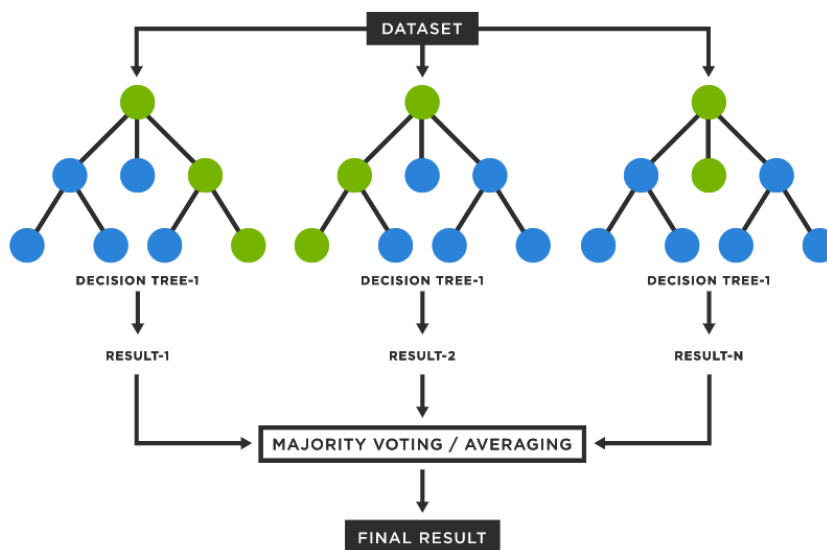


Obrázok 8 Ukážka modelu Decision Tree

### 3.4 Random Forest

(Spracoval: Viet Quoc Le)

Náhodný les, anglický Random Forest sa skladá z mnohých rozhodovacích stromov (Decision tree) preto sa volá les. Tento model sa dá používať na klasifikáciu aj regresiu. Každý strom spraví svoju predpoveď, môžeme to prirovnať k hlasovaniu. Každý strom má jeden hlas. Pri klasifikácii výsledná trieda lesu je tá, ktorá mala najviac hlasov. Regresia nepredpovedá triedu, ale číselne vyjadrený predpoklad, ktorý nemusí byť celé číslo. Výsledný výsledok lesu pri regresii je priemerná hodnota výsledkov stromov.



Obrázok 9 Model Random Forest

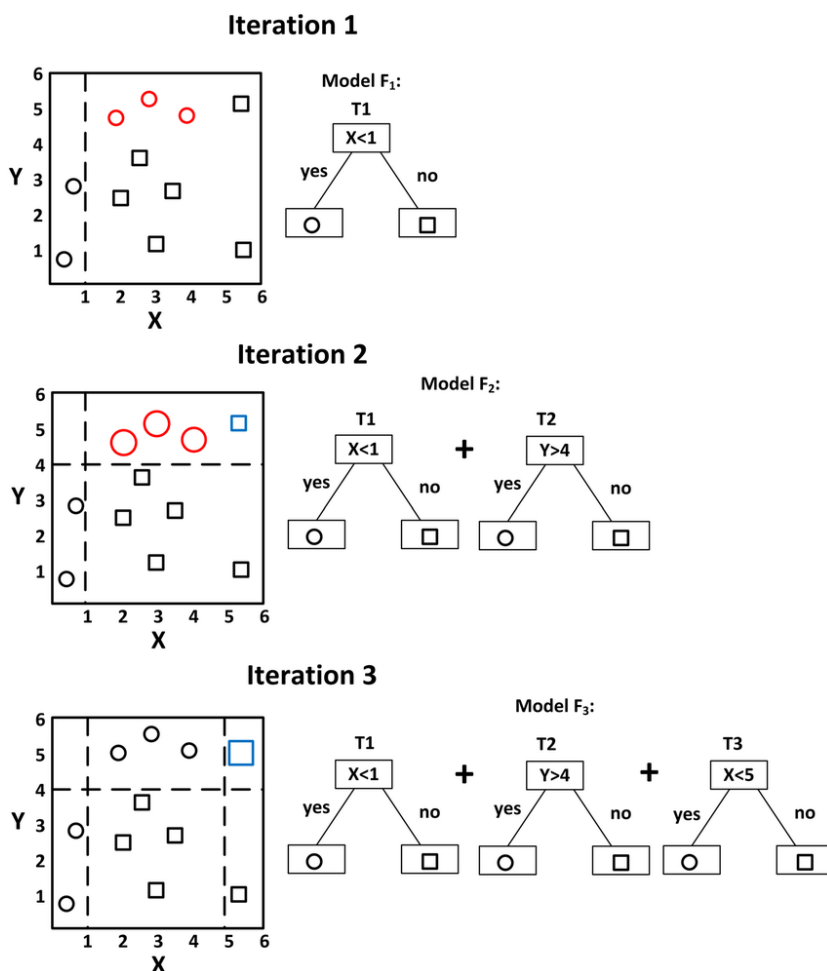
Dôležité pre náhodný lesy je zabezpečiť variáciu a nezávislosť. Toto dosiahneme použitím Bagging metódy. Bagging rozdelí trénovací dataset na menšie náhodne datasety, ktoré sa pridelia stromom, každý strom preto bude mať náhody dataset.

### 3.5 XGBoost

(Spracoval: Filip Frank)

XGBoost (skratka pre Extreme Gradient Boosting) je implementácia gradient boosting metódy súborového strojového učenia s učiteľom vytvorená pre efektívne riešenie klasifikačných a regresných problémov. Gradient boosting využíva ako slabý klasifikátor (regresor) rozhodovacie stromy, ktoré sú jeden za druhým kombinované do silného klasifikátora (regresora) za účelom minimalizácie chybovej (loss) funkcie. Aby sa zabránilo pretrénovaniu silného regresora, využívajú sa metódy regularizácie pre udržanie rozhodovacích stromov na úrovni slabých klasifikátorov (stromy majú obmedzenú hĺbku, počet listov, využívajú sa L1/L2 regularizácie pre váhy listov, atď.). Hlavnou výhodou XGBoost implementácie je jej škálovateľnosť: XGBoost vylepšuje gradient boosting metódu o optimalizačné algoritmy, ktoré zabezpečujú komparatívne rýchlejší priebeh

trénovania a minimalizujú nároky na potrebné výpočtové zdroje. Informácie k implementácii XGBoost sú dostupné na <https://arxiv.org/abs/1603.02754>.



Obrázok 10 Jednoduchá ukážka princípu gradient boosting metódy s využitím rozhodovacích stromov.

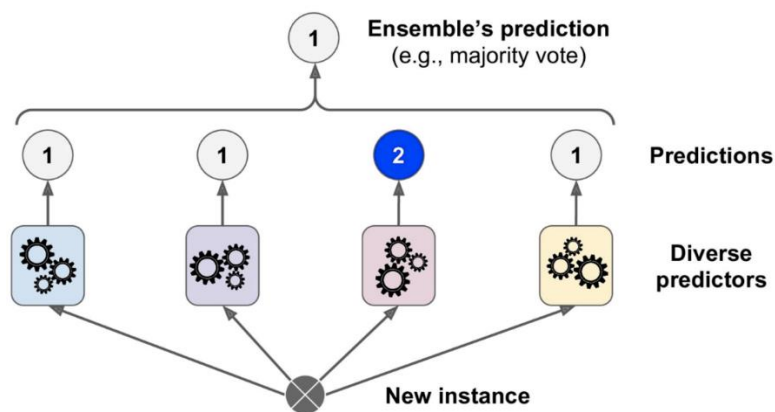
## 3.6 Voting Ensemble

(Spracoval: Tomáš Singhofer)

Model Voting Ensemble je model, ktorý predikuje odpoveď na základe viacerých predpovediach z iných modelov. Môže byť použitý ako na regresiu tak aj na klasifikáciu

V prípade regresie model vytvorí predpoveď, ktorá je priemerom viacerých iných regresných modelov.

Klasifikácia sa delí na dve hlavné metódy a to hard a soft voting. Pri metóde hard voting model sčíta hlasy z viacerých modelov a vyberie sa trieda s najväčším počtom hlasov. Pri soft votingu sa sčítajú pravdepodobnosti a vyberie sa trieda s najväčšou pravdepodobnosťou



Obrázok 11 Voting Ensemble

### 3.7 Stacking Ensemble

(Spracoval: Filip Frank)

Stacking je technika súborového strojového učenia, pre ktorú sú typické dve úrovne tréningu modelu:

1. Trénovanie jednotlivých podmodelov (môžu to byť slabé i silné klasifikátory/regresory)
2. Trénovanie druhej úrovne, v ktorej sa model učí predikcie týchto podmodelov nakombinovať tak, aby poskytol optimálnu odpoveď

### 3.8 Elastic Net

(Spracoval: Branislav Šipula)

Elastic Net je rozšírením lineárnej regresie, ktorá pridáva regulačné sankcie k funkcii straty počas tréningu. Ako vyhodnotiť model elastickej siete a použiť konečný model na predpovedanie nových údajov. Ako nakonfigurovať model Elastic Net pre nový súbor údajov pomocou vyhľadávania v mriežke a automaticky.

Jedným z populárnych trestov je penalizácia modelu na základe súčtu hodnôt štvorcových koeficientov. Toto sa nazýva trest L2. Pokuta L2 minimalizuje veľkosť všetkých koeficientov, aj keď zabráňuje odstráneniu akýchkoľvek koeficientov z modelu.

Ďalšou populárnou penalizáciou je penalizácia modelu na základe súčtu hodnôt absolútnych koeficientov. Toto sa nazýva trest L1. Pokuta L1 minimalizuje veľkosť

všetkých koeficientov a umožňuje, aby sa niektoré koeficienty minimalizovali na hodnotu nula, čím sa z modelu odstráni prediktor.



## 4 Doterajšie výsledky

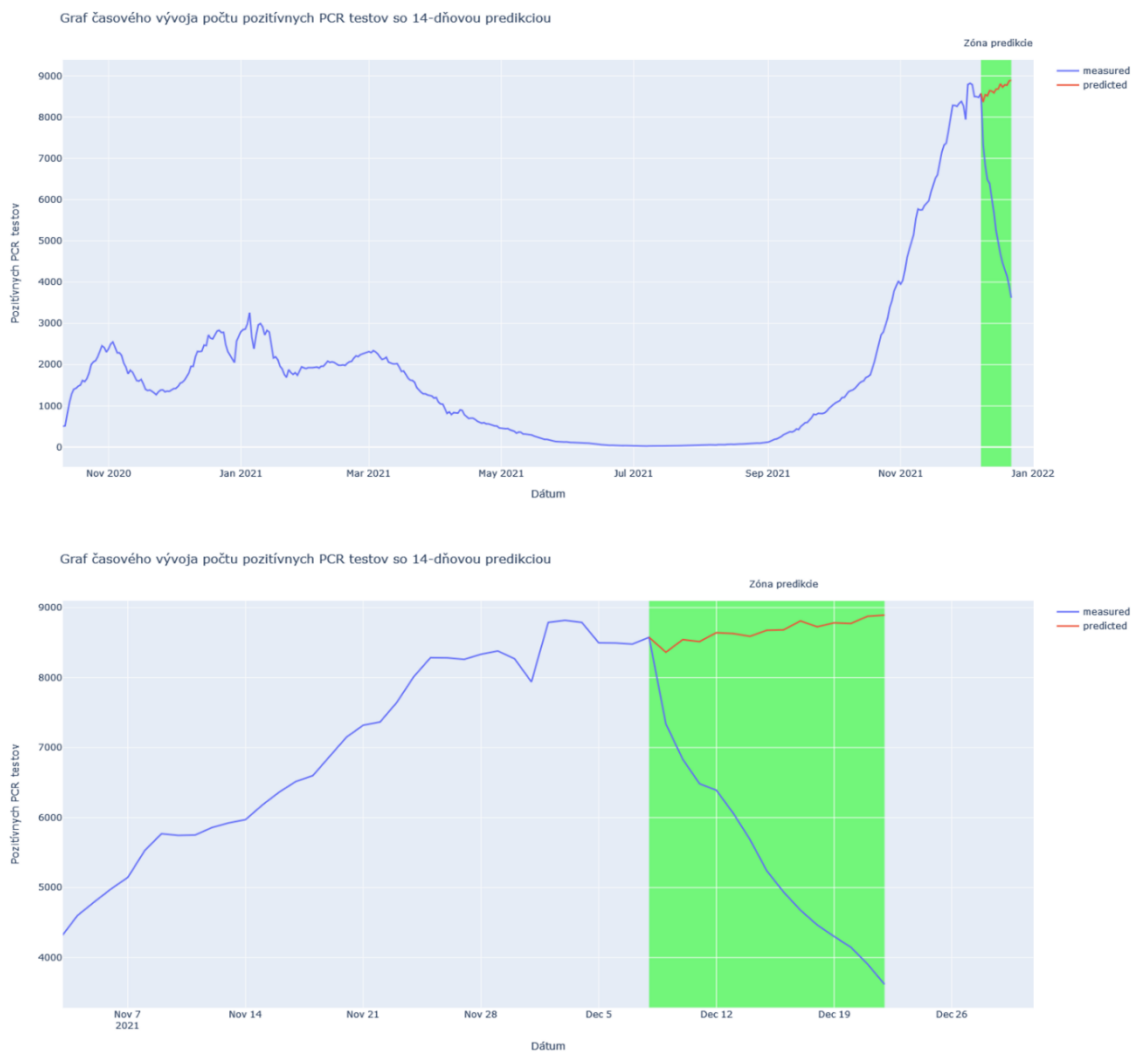
### 4.1 Predikcia pozitívnych PCR testov

(Spracoval: Filip Frank)

Stĺpce datasetu (boli použité pre tréningové predikčných modelov):

1. dátum (10.10.2020 – 8.12.2021)
2. denný počet pozitívnych PCR testov
3. denné percento pozitívnych PCR testov
4. denný počet pozitívnych Ag testov
5. denné percento pozitívnych Ag testov

Víťazný predikčný model: Arimax



Obrázok 12 Graf časového vývoja počtu pozitívnych PCR testov (modrým) so 14-dňovou predikciou (červeným)

Predikcia bola vykonaná na 14 nasledujúcich dní (9.12-22.12.2021), na pôvodných dátach bol aplikovaný algoritmus 7-dňového moving average pre vyhladenie vysvetliteľne nekonzistentných dát (napr. periodické významné poklesy denného počtu testov v dôsledku nízkej miery testovania cez víkendy).

Na grafe vidíme červenou predpoveď denného počtu pozitívnych PCR testov, reálny počet pozitívnych testov reprezentuje modrý graf. Medzi predpovedaným a reálne zisteným počtom PCR testov je výrazný rozdiel, pretože práve v čase, kedy začíname s predikciou (9.12.2021), nastal prudký zlom v počte zistených pozitívnych prípadov. Natrénovaný model nie je tak dokonalý, aby dokázal tento pokles predpovedať, namiesto toho pokračuje v predpovedi podľa trendu predchádzajúcich dní. Výrazný rozdiel pozorujeme už v prvom dni predikcie, kde model predpovedá počet PCR testov 8361.53, pričom reálna hodnota je 7337.57 (pozn.: hodnota po aplikácii 7-dňového moving average). V poslednom dni predikcie model predpovedá 8892.61, reálne však prírastok zodpovedá číslu 3618.14.

## 4.2 Predikcia pozitívnych Ag testov

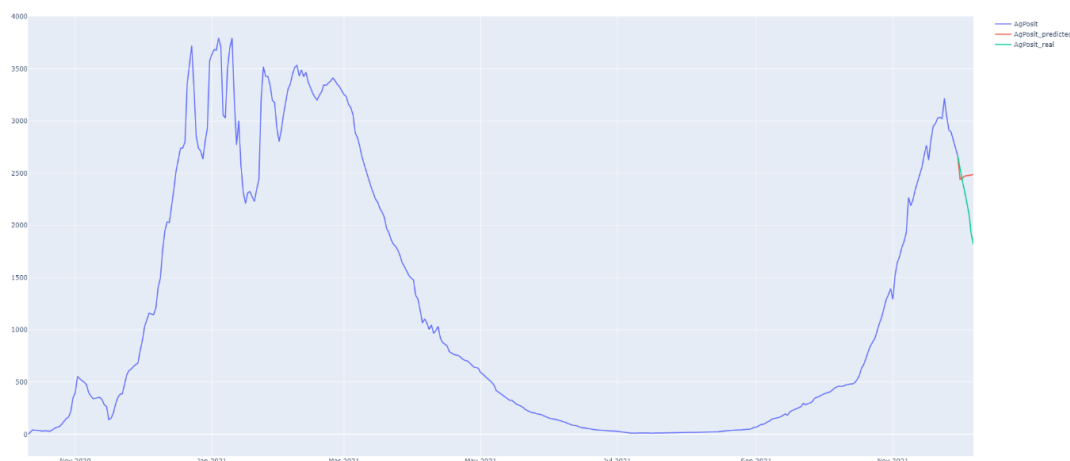
*(Spracoval: Juraj Puszter)*

Model neurónovej siete sme trénovali na základe:

- pozitívnych antigénových testov
- pozitívnych antigénových testov v percentách
- počte pozitívnych PCR testov
- počte pozitívnych PCR testov v percentách

Pri predpovedaní sme dáta pozitívnych antigénových testov nepoužili, nakoľko tie sme predpovedali.

Predpoveď a trénovanie modelu neurónovej siete sa uskutočnilo na dátach zozbieraných z datasetu poskytnutým Inštitútom Zdravotných Analýz, ktorý sa nachádza na githube. V datasete s názvom daily stats sa nachádzajú rôzne údaje, vrátane údajov o celkovom a tiež pozitívnom počte antigénových a PCR testov. Z týchto dát sme vyjadrili počet pozitívnych testov v percentách. Z datasetu sme využili dáta od dňa 11. Októbra 2020 do 30. Novembra 2021. Predikovali sme na 7 dní, od 1. do 7. Decembra 2021. Dáta boli pred predikciou upravené pomocou 7 dňového pohyblivého priemeru. Predikciu sme neskôr porovnali s reálnymi dátami a zobrazili na grafe.



Obrázok 13 Graf predikcie Ag z PCR, PCR%, Ag, Ag%

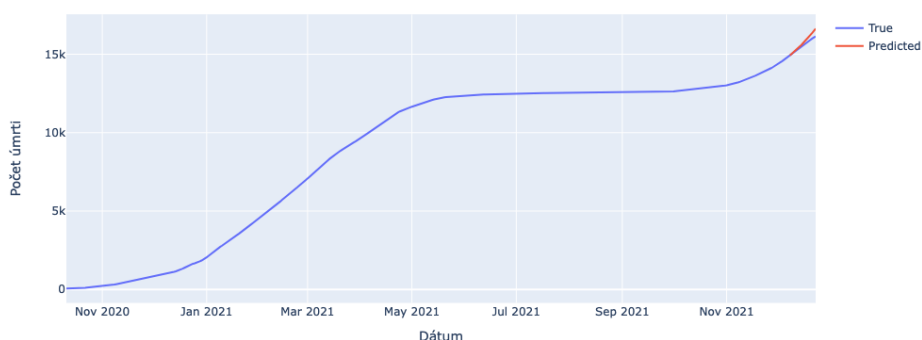
Modrá čiara na grafe znázorňuje počet pozitívnych antigénových testov. Červená čiara znázorňuje našu predikciu a zelená reálne hodnoty. Ako na grafe môžeme vidieť, naša predikcia sa od reálnych hodnôt značne líši. Predpovedaný počet antigénových testov má 7. Decembra hodnotu 2,490.341 a reálna hodnota je 1,817.71. Pri porovnaní predikovaných a reálnych hodnôt pomocou RMSE nám vyšlo 378.21833995587417. Vieme vyčítať, že reálne hodnoty klesli značne viac, ako naše predikované.

### 4.3 Predikcia úmrtí z viacerých ukazovateľov

(Spracoval: Viet Quoc Le)

Semestrálnou úlohou bolo spraviť predikciu počtu úmrtí na 14 dni, z denných štatistík na Slovensku, ktoré sú dostupné v repozitári IZA na githube v priečinku DailyStats. Dáta sa začali zbierať od 6.3.2020, nakoľko dáta boli na začiatku nekonzistentné až po 10.10.2020, kedy sa začalo plošné testovanie začali byť dáta konzistentné, som použil dáta od 10.10.2020. Použité príznaky na tréning boli denne pozitívne Ag, pozitívne Ag v percentách, pozitívne PCR, pozitívne PCR v percentách. Pripravený súbor s údajmi bol vložený do Azure na tréning a vyhodnotenie najlepšieho modelu. Najlepší model bol neskôr použitý na predikciu.

3) Predikcia úmrtí z (úmrtí, PCR, PCR%, Ag, Ag%)



3) Predikcia úmrtí z (úmrtí, PCR, PCR%, Ag, Ag%)



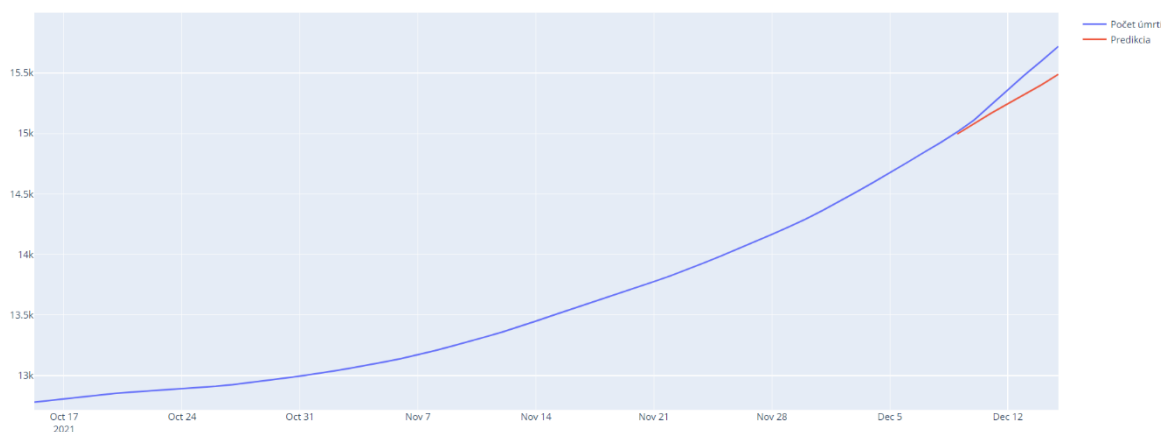
Obrázok 14 Graf predikcie úmrtí z PCR, PCR%, Ag, Ag%

Predpovede modelu spočiatku sú výrazne lepšie ako neskoršie predpovede, po 7 dní sa vzdialenosť medzi predpoveďou a skutočnými údajmi zväčšuje. Predpovedaný počet úmrtí na konci, teda dňa 23.12.2021 je 16 618, skutočný počet úmrtí je 16 124.

## 4.4 Predikcia úmrtí z úmrtí

(Spracoval: Tomáš Singhofer)

Mojím zadáním bolo predikovať celkový počet úmrtí. Sieť sa učila pomocou zozbieraných dát z minulosti z githubu inštitútu zdravotných analýz a na základe toho predikovala vývoj situácie na ďalších 7 dní (9.12. – 15.12.). Sieť predikovala počet úmrtí na deň 15.12. niečo menej ako 15 500. V skutočnosti bol počet úmrtí v tento deň až 15 730. Pri mojom tréovaní mal najlepšie výsledky model Voting Ensemble.

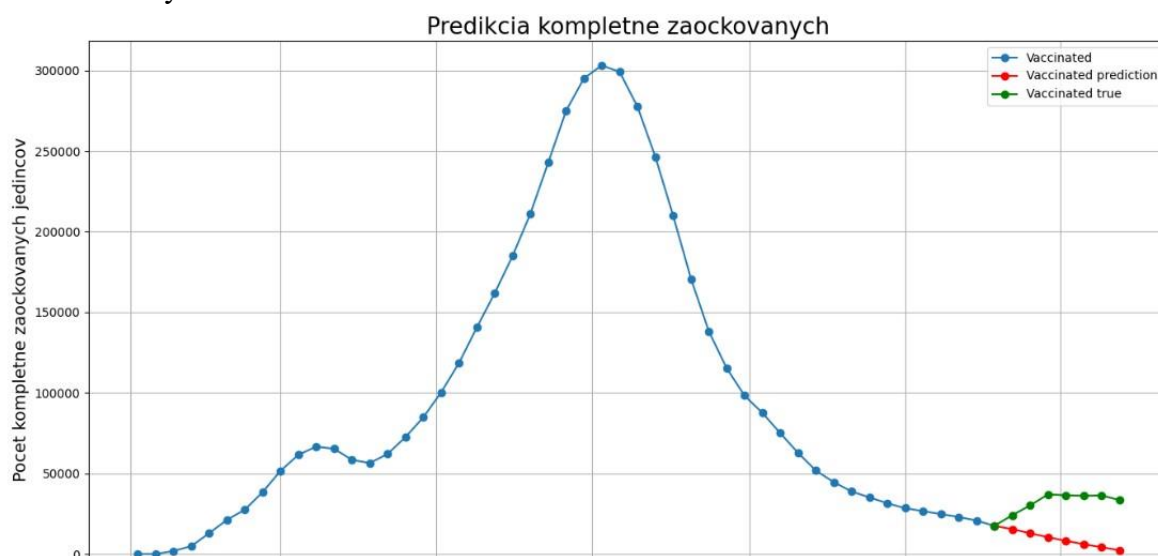


Obrázok 15 Graf predikcie úmrtí z úmrtí

## 4.5 Predikcia kompletne zaočkovaných

(Spracoval: Marek Kačmár)

Mojou finálnou úlohou bolo predikovať vývoj počtu pribúdajúcich plne zaočkovaných jedincov, ktorým bola podaná 2. dávka. Prvým krokom bolo pripraviť si dáta, ktoré budem predikovať. Zvolil som si nasledovné [csv](#) z github inštitútu zdravotných analýz. Po úprave sa mi podarilo zoradiť dáta podľa jednotlivých týždňov a počtu druhých dávok, ktoré indikovali o plne zaočkovanom jedincovi. Na pripravené dáta som aplikoval moving average so 7-dňovým oknom, následne som z Azure použil model, ktorý najlepšie predikoval môj dataset, a ten som využil vo svojej vizualizácii predpovede vývoja počtu plne zaočkovaných na ďalších 7 týždňov.



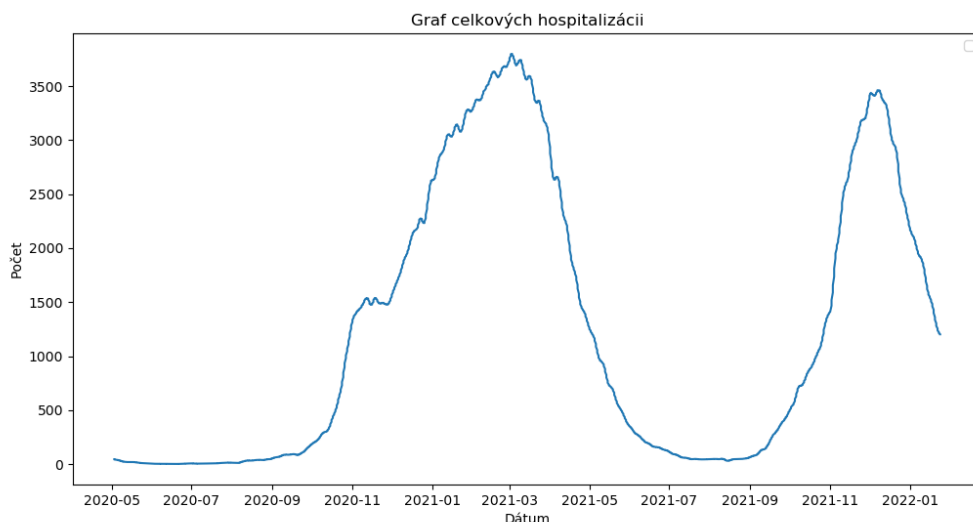
Obrázok 16 Graf predikcie kompletne zaočkovaných

Na grafe môžeme pozorovať porovnanie predikovaného vývoja počtu plne očkovaných a skutočného vývoja počtu plne zaočkovaných jedincov v jednotlivých týždňoch. Predikcia bola vytvorená ešte v polovici decembra a samotné porovnanie bolo vykonané 22. januára. Ako môžeme vidieť na grafe, počet pribúdajúcich plne zaočkovaných ľudí v jednotlivých týždňoch klesal postupne k nule, od čoho sa odvíjala aj jeho predpoveď, avšak vplyvom rôznych činiteľov, ktoré v predpovedi neboli zahrnuté, ako napr. zvýšenie počtu nakazených a hospitalizovaných (strach z nakazenia/úmrtia), politická podpora zaočkovanosti a i. tento počet zvýšili, čo je možné vidieť na skutočnom počte plne zaočkovaných v prislúchajúcich týždňoch.

## 4.6 Graf celkových hospitalizácií

(Spracoval: Branislav Šipula)

Graf znázorňuje stav/vývoj počtu hospitalizovaných pacientov v nemocniciach za obdobie od 30.04.2020 do 9.12.2021. Z grafu je vidieť, že najvyšší nárast hospitalizovaných pacientov bol v mesiaci marec a druhy najvyšší nárast bol v mesiaci december 2021. Najnižší pokles bol medzi týmito extrémami bolo v letných mesiacoch a z toho najmenší v mesiaci august 2021.



Obrázok 17 Graf celkových hospitalizácií

## 4.7 Graf celkových UPV

(Spracoval: Adam Vozár)

Záverečná úloha pozostávala z vytvorenia grafu obsadenosti miest na umelej pľúcnej ventilácii. K vytvoreniu grafu bolo potrebné použiť dáta z Inštitútu Zdravotných Analýz,



Obrázok 18 Graf časového vývoja UPV

konkrétne súbor `Hospital_UPV_AdmissionDischarge.csv`, ktorý v prvom stĺpci obsahuje dátum a v druhom stĺpci číslo, ktoré predstavuje o koľko sa počet miest na UPV zvýšil resp. znížil. Dataset som doplnil o ďalší stĺpec suma, v ktorom sa počet obsadených miest na UPV kumulatívne sčítuje ku každému dátumu. Z grafu môžeme vidieť viditeľný nárast UPV od októbra 2020 ktorý dosiahol svoje maximum dňa 20. marca 2021 s počtom 422 UPV.

# Zoznam použitej literatúry

## Azure

<https://searchcloudcomputing.techtarget.com/definition/Windows-Azure>

<https://ccbtechnology.com/what-microsoft-azure-is-and-why-it-matters/>

<https://docs.microsoft.com/en-us/azure/machine-learning/overview-what-is-azure-machine-learning>

<https://docs.microsoft.com/en-us/azure/machine-learning/overview-what-is-machine-learning-studio>

<https://docs.microsoft.com/en-us/azure/machine-learning/concept-automated-ml#when-to-use-automl-classification-regression-forecasting--computer-vision>

<https://caiomsouza.medium.com/aml-azure-machine-learning-introduction-13505d3393b9>

## Seasonal Naive

[https://en.wikipedia.org/wiki/Forecasting#Na%C3%AFve\\_approach](https://en.wikipedia.org/wiki/Forecasting#Na%C3%AFve_approach)

<https://otexts.com/fpp2/simple-methods.html>

<https://medium.com/@sangarshananveera/time-series-forecasting-part-1-d523bcc3acbf>

## ARIMA a ARIMAX

<https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/>

<https://www.r-bloggers.com/2017/05/forecasting-arimax-model-exercises-part-5/>

<https://timeseriesreasoning.com/contents/regression-with-arima-errors-model/>

<https://pyflux.readthedocs.io/en/latest/arima.html>

<https://pyflux.readthedocs.io/en/latest/arimax.html>

## XGBoost

<https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/>

<https://xgboost.readthedocs.io/en/stable/tutorials/model.html>

<https://arxiv.org/abs/1603.02754>

[https://www.researchgate.net/figure/A-simple-example-of-visualizing-gradient-boosting\\_fig5\\_326379229](https://www.researchgate.net/figure/A-simple-example-of-visualizing-gradient-boosting_fig5_326379229)

## Random Forest



<https://www.youtube.com/watch?v=cIbj0WuK41w>

<https://course.elementsofai.com/sk/4/3>

<https://builtin.com/data-science/random-forest-algorithm>

Elastic Net

<https://machinelearningmastery.com/elastic-net-regression-in-python/>

Stacking Ensemble

<https://machinelearningmastery.com/stacking-ensemble-for-deep-learning-neural-networks/>