

# Praćenje koncentracije peludi: pristup pomoću rekurentnih neuronskih mreža

---

Marinković, Željana

Undergraduate thesis / Završni rad

2023

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **Josip Juraj Strossmayer University of Osijek, School of Applied Mathematics and Informatics / Sveučilište Josipa Jurja Strossmayera u Osijeku, Fakultet primijenjene matematike i informatike**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:126:079355>

*Rights / Prava:* [In copyright](#)/Zaštićeno autorskim pravom.

*Download date / Datum preuzimanja:* **2024-07-12**



*Repository / Repozitorij:*

[Repository of School of Applied Mathematics and Computer Science](#)





SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU  
FAKULTET PRIMIJENJENE MATEMATIKE I INFORMATIKE

Sveučilišni prijediplomski studij Matematika i računarstvo

# Praćenje koncentracije peludi: pristup pomoću rekurentnih neuronskih mreža

ZAVRŠNI RAD

Mentor:

**izv. prof. dr. sc.  
Domagoj Matijević**

Kandidat:

**Željana Marinković**

Osijek, 2023.

## Sažetak

U ovome radu bavimo se predikcijom koncentracije triju vrsta peludi u zraku. Budući da podatke koristimo u obliku vremenskih nizova, rješavanju problema pristupit ćemo pomoću rekurentnih neuronskih mreža. Na početku uvodimo pojam rekurentnih neuronskih mreža te navodimo neka njihova ograničenja. Zatim, kako bismo prevladali ta ograničenja, uvodimo podvrstu RNN-a, LSTM, koji koristimo za izradu modela. Nadalje, prezentiramo korištenu bazu podataka i varijable o kojima ovisi koncentracija peludi te način na koji dijelimo te podatke u svrhu treniranja modela. Potom opisujemo arhitekturu modela, prikazujemo dobivene rezultate te komentiramo stvarne koncentracije peludi sada i u prošlosti.

## Ključne riječi

strojno učenje, neuronske mreže, rekurentne neuronske mreže, LSTM, koncentracija peludi

# Pollen concentration monitoring: an approach using recurrent neural networks.

## Abstract

In this paper, we focus on predicting the concentration of three types of pollen in the air. Since we are working with data in the form of time series, we approach the problem using recurrent neural networks. Initially, we introduce the concept of recurrent neural networks and mention some of their limitations. To overcome these limitations, we introduce a subtype of RNN, LSTM, which we use to build our model. Furthermore, we present the dataset used and the variables that affect pollen concentration, as well as how we divide this data for training the model. Next, we describe the model's architecture, display the obtained results and provide information on actual pollen concentrations, both present and in the past.

## Keywords

machine learning, neural networks, recurrent neural networks, LSTM, pollen concentration

# Sadržaj

Uvod	1
1 Rekurentna neuronska mreža - RNN	2
2 LSTM - Long Short-Term Memory	3
3 Obrada podataka	5
3.1 Transformacija vremenskih varijabli . . . . .	5
3.2 Razdvajanje podataka . . . . .	6
4 Model	7
4.1 Funkcija gubitka . . . . .	7
4.2 Arhitektura modela . . . . .	7
5 Rezultati	8
5.1 Pelud breze . . . . .	9
5.2 Pelud trave . . . . .	12
5.3 Pelud ambrozije . . . . .	13
6 Zaključak	17
Literatura	18

## Uvod

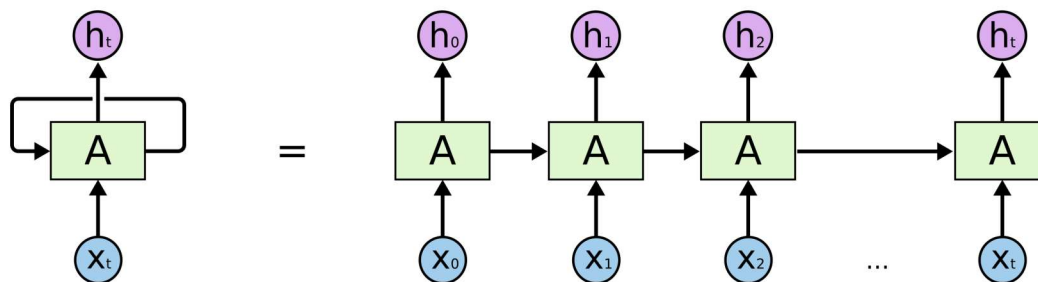
Ambrozija, breza i trava biljka su koje ponajviše rastu na zapuštenim zemljištima, livadama, šumama, uz putove, polja, među usjevima, vrtovima, ali i svuda oko nas. Među biljakama su koje proizvode najveću količinu peludi. Budući da peludna alergija (prekognjerna reakcija imunološkog sustava na inače bezopasne tvari iz okoliša) zahvaća između 10% i 30% svjetske populacije, cilj ovog završnog rada bio je napraviti model koji će predviđati koncentraciju ovih triju vrsta peludi u zraku.

Budući da podaci o koncentraciji peludi nisu nezavisni, tj. ona ovisi o koncentraciji, klimatskim faktorima i elementima od prethodnih dana, klasična feedforward neuronska mreža ne bi mogla dati dovoljno dobre rezultate. Stoga je u izradi ovoga rada korištena vrsta rekurentnih neuronskih mreža – LSTM (long short-term memory).

Također, korištena je baza podataka "real\_for\_all\_NS\_2000\_2021.csv" u kojoj su prikupljeni podaci za razdoblje 2000.-2021. na području Novoga Sada.

# 1 Rekurentna neuronska mreža - RNN

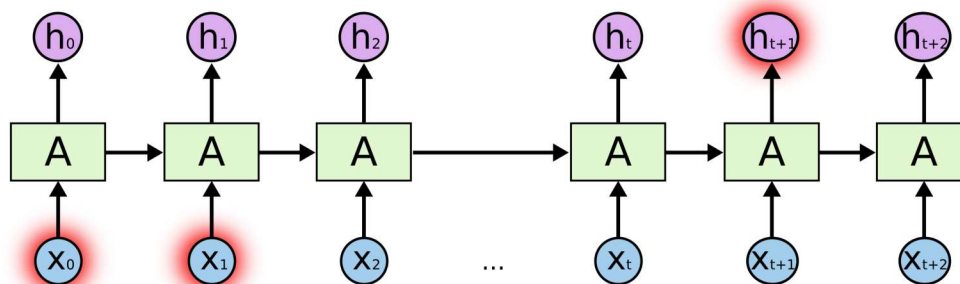
Rekurentne neuronske mreže vrsta su neuronskih mreža koje kao ulaz primaju niz podataka, a taj niz može biti varijabilne duljine. Za razliku od ostalih neuronskih mreža, u rekurentnim se mrežama pojavljuju petlje unutar modela (Slika 1.<sup>1</sup>) što omogućava očuvanje informacija o prethodnim podacima iz predanog niza. Te informacije pohranjuju se u skriveno stanje (eng. hidden state). Za stvaranje sljedećeg skrivenog stanja koristi se sljedeći podatak iz niza te prethodno skriveno stanje. Početno skriveno stanje uglavnom odgovara nulvektoru. Zbog navedenih značajki, rekurentne neuronske mreže koriste se pri obradi prirodnog govora, teksta, vremenske prognoze te općenito podataka koji ovise o vremenu.



Slika 1: Prikaz petlje u RNN-u kao niza kopija iste mreže radi lakšeg shvaćanja

RNN mreže mogu se konstruirati na više načina, ovisno o problemu koji se rješava. Tako se može konstruirati RNN koja pri obradi svakoga podatka daje izlaz, kao i ona koja daje izlaz tek nakon što je obrađen čitav niz podataka.

Problemi koji nastaju kod korištenja RNN-a su nestajući (eng. vanishing) i eksplodirajući (eng. exploding) gradijent koji označavaju da je gradijent pri treniranju modela počeo težiti u 0 ili  $\infty$ . Do toga može doći kada ulazni niz sadrži velik broj podataka te postoji povezanost među njima, no i „velika udaljenost“ pa se informacija o povezanosti kroz iteracije „izgubi“ (Slika 2.<sup>2</sup>).



Slika 2: Problem kod udaljenih elemenata niza

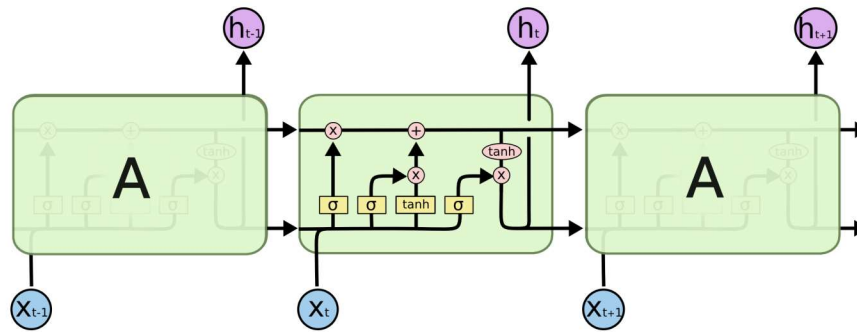
Neka od ponuđenih rješenja za ovaj problem su podvrste rekurentnih neuronskih mreža, LSTM (Long Short-Term Memory) i GRU (Gated Recurrent Unit). Za izradu ovoga rada korišten je LSTM model.

<sup>1</sup>Preuzeto s <https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e>

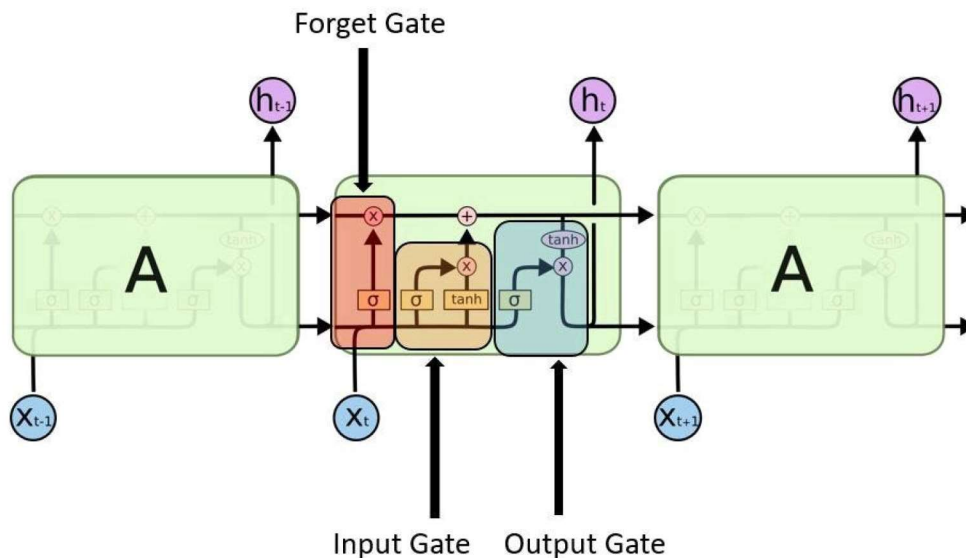
<sup>2</sup>Preuzeto s <https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e>

## 2 LSTM - Long Short-Term Memory

LSTM vrsta je rekurentne neuronske mreže osmišljena s ciljem sprječavanja nestajućeg gradijenta. Kako bi se omogućilo dugoročno pohranjivanje informacija iz prethodnih stanja uvode se pojmovi ćelije (Slika 3.<sup>3</sup>)– skupa operacija kojim neuronska mreža pomoću ulaza i prijašnjeg skrivenog stanja stvara novo skriveno stanje te stanja ćelije - dodatnog skupa parametara koji može pohranjivati važne informacije iz prethodnih koraka, a kontroliraju ga ulazna, izlazna i vrata zaboravljanja(Slika 4.<sup>4</sup>).



Slika 3: Struktura ćelije



Slika 4: Ulazna, izlazna i vrata zaboravljanja

Budući da često promatramo varijable koje ovise o vremenu, umjesto  $i$  - tog podatka niza govorit ćemo koraku u trenutku  $t$  ili samo koraku  $t$ .

Pokazat ćemo kako se računaju pojedina stanja u koraku  $t$ :

**Ulazna vrata** ( $i_t$ ) - (eng. input gates) :  $i_t = \sigma(W_i \cdot [H_{t-1}, X_t] + b_i)$

<sup>3</sup>Preuzeto s <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

<sup>4</sup>Preuzeto s <https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e>



**Izlazna vrata** ( $o_t$ ) - (eng. output gates) :  $o_t = \sigma(W_o \cdot [H_{t-1}, X_t] + b_o)$

**Vrata zaboravljanja** ( $f_t$ ) - (eng. forget gates) :  $f_t = \sigma(W_f \cdot [H_{t-1}, X_t] + b_f)$

Treba napomenuti da uporaba  $\sigma$  (sigmoid) funkcije u ovom problemu ima smisla jer ona vraća broj iz intervala  $[0, 1]$  što možemo shvatit kao vjerojatnost da neku informaciju zadržimo u ćeliji, propustimo ili "zaboravimo".

**Stanje ćelije** ( $C_t$ ) :  $\tilde{C}_t = \tanh(W_c \cdot [H_{t-1}, X_t] + b_c)$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

**Skriveno stanje** ( $h_t$ ) :  $H_t = o_t \odot \tanh(C_t)$

Oznaka  $\odot$  predstavlja produkt po elementima,  $\tanh$  aktivacijsku funkciju *tangens hiperbolni*, a  $[H_{t-1}, X_t]$  je konkatencija prethodnog skrivenog stanja  $H_{t-1}$  i koraka  $X_t$ . Parametri  $W_i, W_f, W_c, W_o, b_i, b_f, b_c, b_o$  su dobiveni tijekom procesa učenja.

### 3 Obrada podataka

Kao što je već navedeno, za izadu završnoga rada korištena je baza podataka "real\_for\_all\_NS\_2000" sa podacima iz razdoblja 2000.-2021. na području Novoga Sada. Navest ćemo koje su sve varijable korištene:

**RA** - je li toga dana padala kiša(1 - da, 0 - ne)

**SN** - snijeg(0/1)

**TS** - olujno nevrijeme(0/1)

**FG** - magla(0/1)

**MBV** - maksimalna brzina vjetra toga dana

**SBV** - srednja brzina vjetra

**PAD** - padaline (u milimetrima)

**VLZ** - vlažnost zraka

**ATT** - atmosferski tlak

**MNT** - minimalna dnevna temperatura

**MKT** - maksimalna dnevna temperatura

**SRT** - srednja dnevna temperatura

**GOD** - godina

**MSC** - mjesec

**DAN** - dan u mjesecu

**RBD** - redni broj dana u godini

**LOK** - lokacija

**PRAM** - prosječna koncentracija ambrozije

**PRBR** - prosječna koncentracija breze

**PRTR** - prosječna koncentracija trave

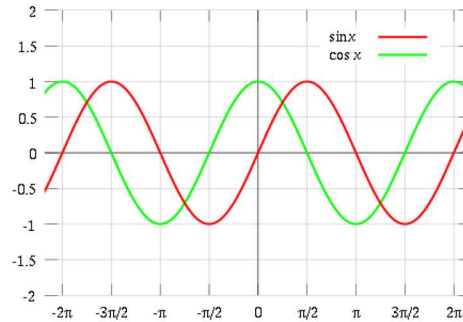
Pri korištenju podataka za potrebe ovoga modela nije korištena varijabla LOK budući da smo imali podatke samo s lokacije Novi Sad. Važno je napomenuti da se ona, kada bismo imali podatke s više lokacija, nikako ne bi smjela ukloniti, no sada nismo htjeli modelu dodavati parametre koji nam neće biti korisni. Također, provođenjem `cor.test()` testova u R-studiju( [4]) i dobivanju p-vrijednosti veće od 0.05 , zaključili smo da ne možemo tvrditi niti jedna vrsta peludi u danoj bazi ovisi atmosferskom tlaku pa ni ta varijabla nije korištena( u slučaju korištenja podataka s neke druge lokacije, trebalo bi ponovno provesti te testove na podacima s nove lokacije kako bi se utvrdilo postoji li korelacije između te varijable i koncentracija peludi).

#### 3.1 Transformacija vremenskih varijabli

Budući da želimo predvidjeti koncentraciju peludi u zraku, a koristimo LSTM model, modelu ćemo predavati nizove od T dana. Pri tome se javlja pitanje kako će naš model znati da nakon prosinca dolazi siječanj, a nakon posljednjeg dana u mjesecu/godini ponovno prvi? Kako ne bi došlo do "zabune" pri računanju u modelu kada su u pitanju takvi prijelazi, napraviti ćemo jednu transformaciju na vrijednostima tih varijabli. Možemo reći da su dani u mjesecu/godini ili mjeseci u godini periodične funkcije, a također su i periodične funkcije *sinus* i *cosinus* (Slika 5.<sup>5</sup>).

Transformacija za mjesece u godini izgleda ovako:

<sup>5</sup>Preuzeto s <https://climatescienceteaching.org/tool/trigonometry>



Slika 5: Periodični graf sinusa i kosinusa

$$\text{sinusMSC} := \sin(2 * \pi * \text{MSC}/12)$$

pri čemu je `sinusMSC` nova varijabla koja se dobije tako da funkcijom *sinus* djelujemo na kvocjent umnoška broja 2,  $\pi$  i pripadne vrijednosti varijable `MSC` s ukupnim brojem mjeseci u godini([5]).

## 3.2 Razdvajanje podataka

Budući da želimo napraviti model koji bi nam mogao pomoći u predviđanju koncentracije peludi u zraku u budućnosti, a i podatke grupiramo u vremenske nizove, nikako nema smisla dijeliti ih nasumično. Htjeli bismo da naš model što bolje opisuje najnovije podatke pa smo skup podataka podijelili na sljedeći način - prvih 15 godina služilo nam je kao skup za treniranje, sljedeće tri kao skup za validaciju, a posljednje tri bile su skup za testiranje. Za izradu modela<sup>6</sup>, kao i za obradu podataka korišten je programski jezik Python te biblioteka TensorFlow([8]).

---

<sup>6</sup>Git repozitorij dostupan je na [12]

## 4 Model

### 4.1 Funkcija gubitka

Za učenje modela potrebno je odabrati odgovarajuću funkciju gubitka (eng. loss function). Cilj nam, je kao i inače kod učenja modela, minimizirati funkciju gubitka. Budući da želimo da je razlika između stvarnih podataka i onih koje računa model što manja, odabrana je funkcija  $MSE$  (eng. mean squared error) čiji je općeniti oblik opisan sljedećim izrazom:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

pri čemu je  $y_i$  stvarna vrijednost  $i$ -tog podatka,  $\hat{y}_i$  vrijednost koju je model predvidio za taj podatak, a  $n$  ukupan broj podataka u bazi. Budući da je u našem modelu  $y_i$  3-dimenzionalni vektor, naša funkcija izgleda ovako:

$$MSE = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{3} \sum_{j=1}^3 (\hat{y}_{ij} - y_{ij})^2 \right)$$

pri čemu unutarnja suma zbraja razlike po komponentama vektora, što je bilo potrebno napraviti budući da MSE funkcija poprima skalarne vrijednosti, a ne vektorske.

Moguće je bilo koristiti i neke druge funkcije ([6]), koje su čak za pojedinu vrstu peludi ostvarivale i bolja predviđanja vrijednosti, no događalo se to da je jedna vrsta dosta lošije predviđena pa smo to izbjegli jer želimo da sve koncentracije budu podjednako dobro predviđene

### 4.2 Arhitektura modela

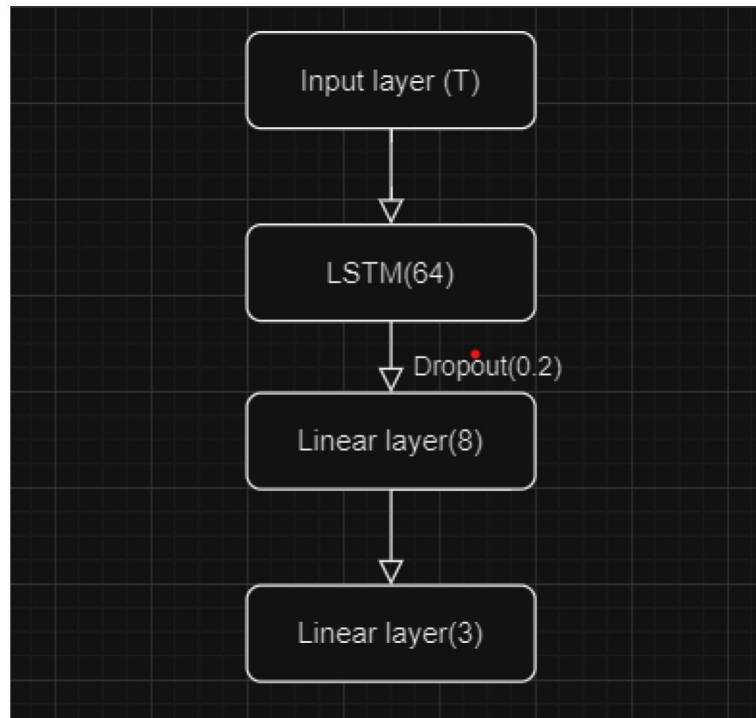
Sljedeći podaci opisuju arhitekturu korištenog modela (moguće vidjeti na slici 6.<sup>7</sup>):

- Ulazni sloj (eng. input layer) - sastoji se od niza od  $T$  dana. Za dobivanje rezultata u ovom radu koristio se  $T = 5$ .
- LSTM - jednosmjerni LSTM sloj sa 64 skrivena stanja.
- Linearni sloj - korištena aktivacijska funkcija je ReLU, a sloj se sastoji od 8 neurona
- Linearni sloj - posljednji sloj sa 3 neurona u kojem je aktivacijska funkcija linearna, a iz kojeg zapravo isčitavamo predviđene koncentracije u zraku za tri vrste peludi.

Također, između LSTM i linearnoga sloja korišten je dropout=0.2, odnosno s vjerojatnošću 0.2 neki su elementi između ova dva sloja postavljeni na nulu kako ne bi došlo do prenaučnosti modela. Za funkciju gubitka odabrali smo ranije navedenu MSE funkciju, a minimizaciju te funkcije dodatno smo ubrzali pomoću optimizatora Adam([7]).

Model smo trenirali kroz 10 epoha - deset smo puta prošli kroz skup podataka za učenje. Kako bismo bili sigurni da će model na kraju izvođenja imati spremljene najbolje parametre te da kroz epohe neće postati prenaučan, uveli smo callback funkciju koja sprema najbolje opažene parametre u modelu, tj. one parametre za koje je na kraju izvođenja epohe greška na skupu za validaciju najmanja.

<sup>7</sup>Izrađeno pomoću online alata dostupnih na <https://app.diagrams.net/>



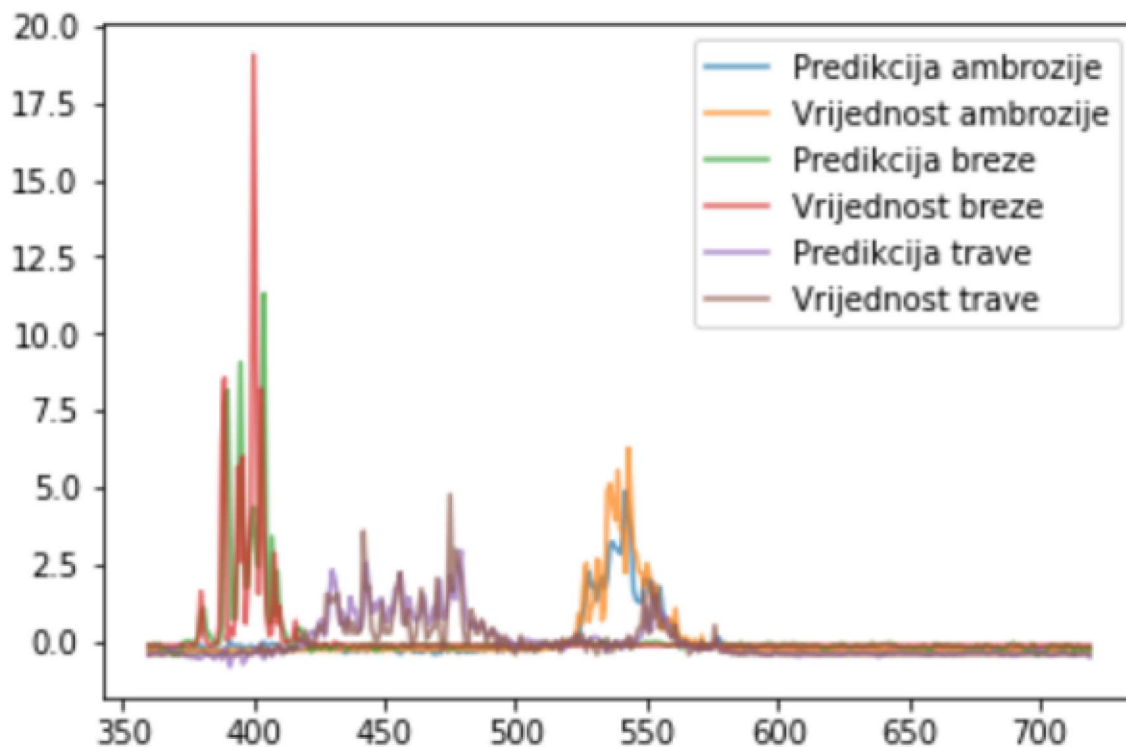
Slika 6: Arhitektura modela

## 5 Rezultati

Model je za  $T = 5$  radio tako da na ulaznom sloju primi niz od 5 dana te vrati predikciju za šesti dan (tj. za "T plus prvi dan"). Kako bismo dobili približnu predodžbu o tome u kojem se dijelu godine pojavljuje koja vrsta peludi, prikazat ćemo zajednički graf stvarnih koncentracija te onih dobivenih modelom za 2020.-u godinu (Slika 7.). Na legendi je moguće vidjeti kojim su bojama prikazane stvarne vrijednosti i predikcije za pojedine vrste peludi.

Budući da je graf nepregledan i ovdje je zbog lakše predodžbe raspodjele peludi kroz godinu, sada ćemo za svaku pojedinu pelud prikazati vrijednosti koje smo dobili i usporediti ih sa stvarnim vrijednostima. (Napomena - radi bolje preglednosti na grafovima će biti prikazani samo mjeseci u kojima se ta vrsta peludi pojavljuje jer u ostalim mjesecima neke uopće nisu prisutne te u tim mjesecima nije došlo do odstupanja predviđenih vrijednosti od stvarnih ). Također, podaci na grafovima normalizirani su pomoću normalne distribucije s očekivanjem i varijancom koncentracija peludi jer zbog stršećih vrijednosti ne možemo podatke pokazati dovoljno dobro kada oni sadrže stvarne vrijednosti.

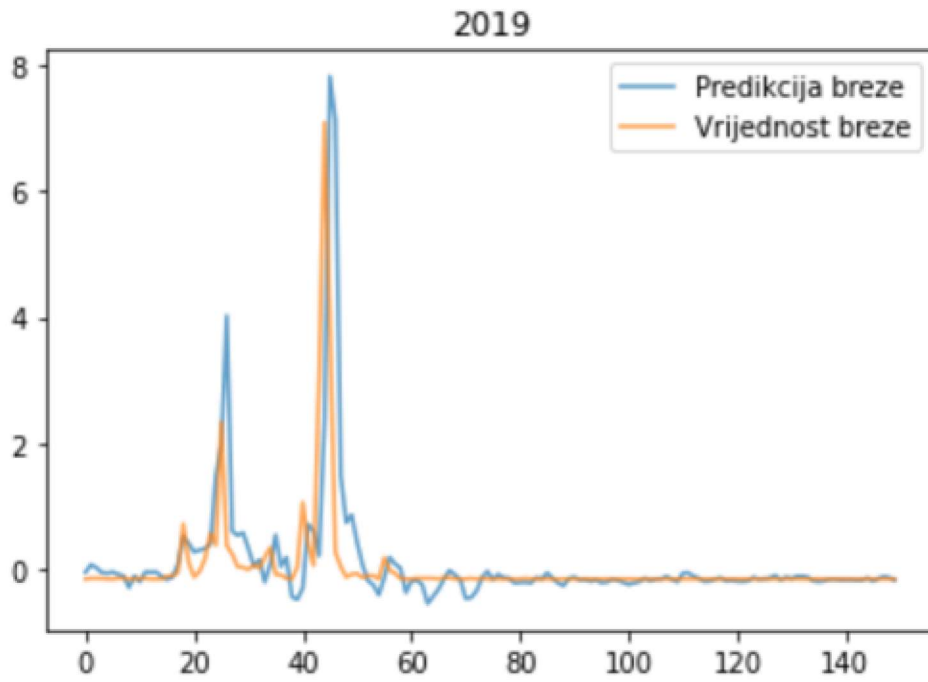




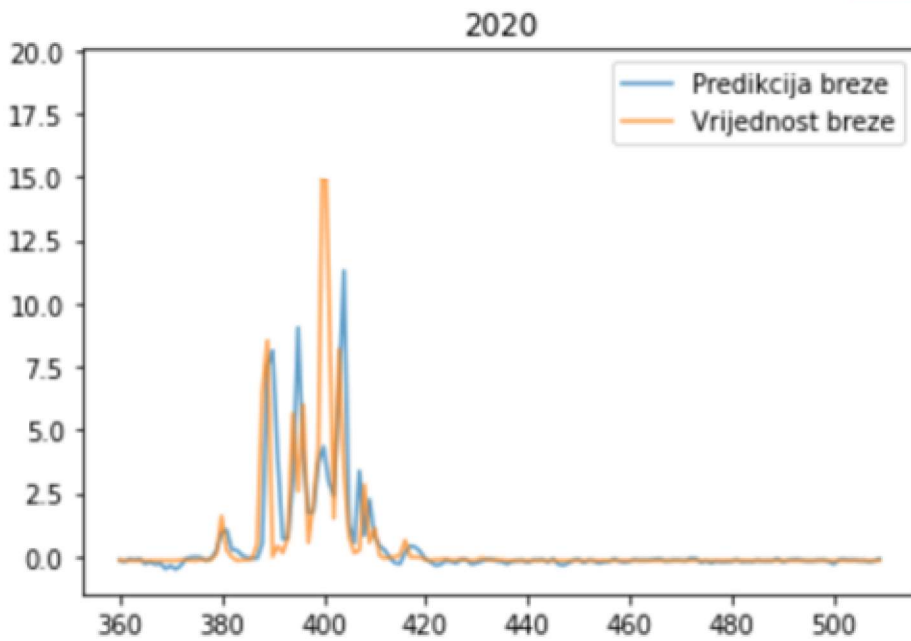
Slika 7: Predikcije i stvarne koncentracije triju vrsta peludi u 2020.godini

## 5.1 Pelud breze

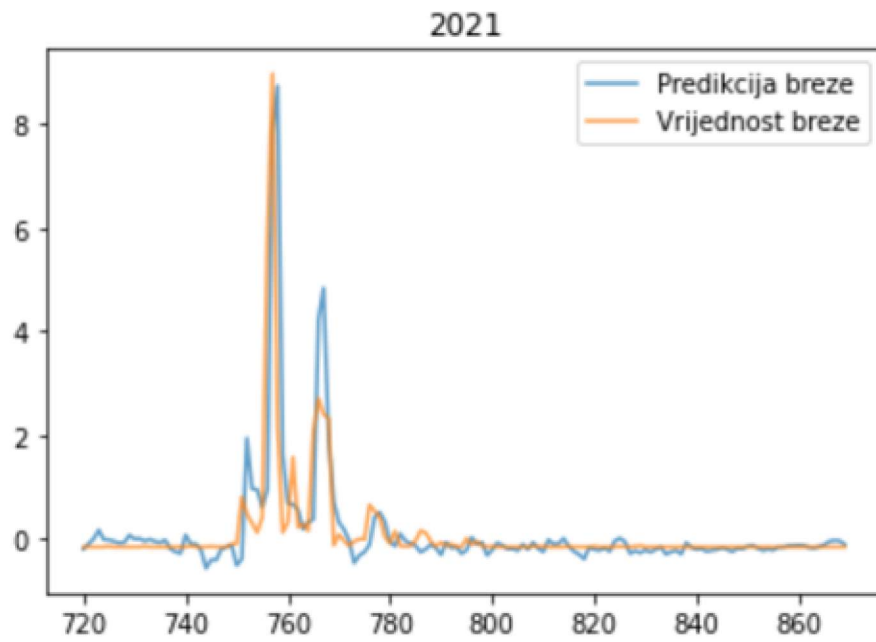
Breza je prilagodljiva i vrlo rasprostranjena vrsta drveća, a često se sadi kao ukrasna drvenasta biljka u gradovima. U skupini drveća pelud breze najjači je alergen te se, uz pelud ambrozije, smatra najjačim peludnim alergenom ([9]). Cvat breze počinje krajem ožujka i početkom travnja. Brzo doseže visoke koncentracije u zraku, a kasnije se u nižim koncentracijama zadržava do kraja svibnja. Na slikama 8., 9. i 10. prikazane su predikcije (plava boja) i stvarne (narančasta boja) vrijednosti breze za godine 2019., 2020. i 2021. Kao što je navedeno, koncentracije u zraku prvo su izuzetno niske, kasnije u vrijeme cvata brzo dosežu vrlo visoke koncentracije u zraku koje na kraju cvata breze naglo opadaju. Pri velikom rastu koncentracije možemo vidjeti da postoje odstupanja stvarnih vrijednosti i predikcija, no možemo zamijetiti da predikcije ipak prate „oblik“ grafa stvarnih vrijednosti, odnosno da rastu kada rastu i stvarne koncentracije te padaju kada dolazi do pada stvarnih vrijednosti. Stoga, iako predikcije nisu u potpunosti točne, model ima određeni značaj pri pomoći ljudima s alergijama jer za njih u potpunosti točne vrijednosti nisu toliko važne koliko je važno znati da će doći do značajnog rasta ili pada peludi u zraku.



Slika 8: Predikcije i stvarne koncentracije peludi breze u 2019.godini



Slika 9: Predikcije i stvarne koncentracije peludi breze u 2020.godini



Slika 10: Predikcije i stvarne koncentracije peludi breze u 2021.godini

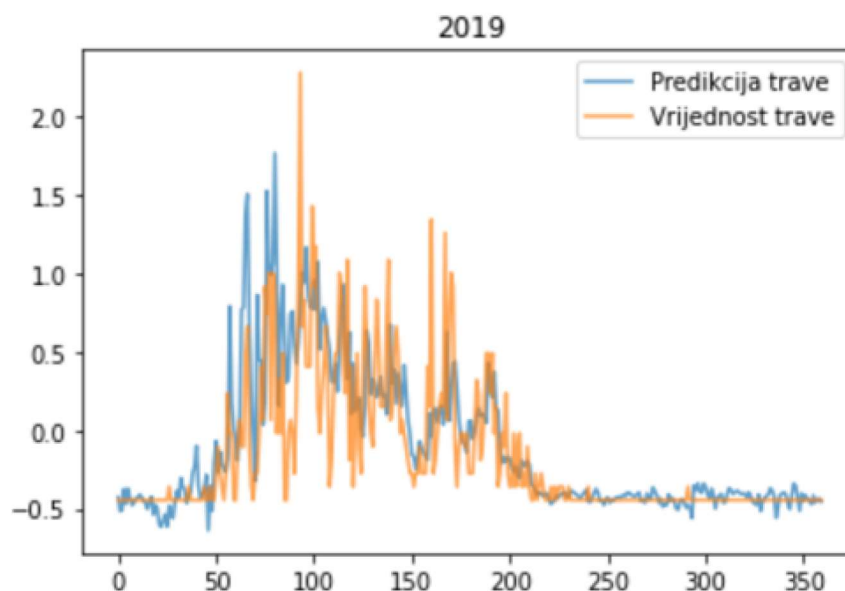


## 5.2 Pelud trave

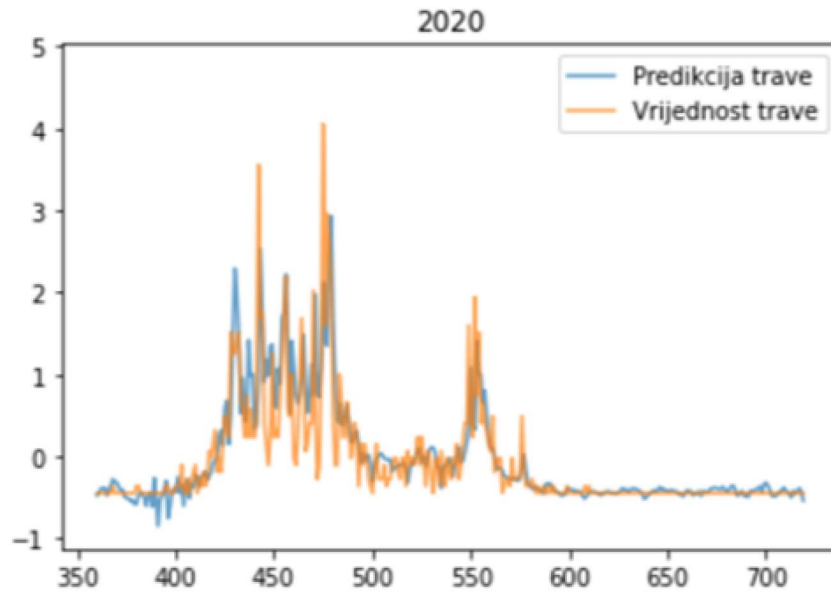
Trave su naziv porodice zeljanica koja broji 740 rodova biljaka. Oko 90% ljudi koji pate od alergija na pelud biljaka alergično je na pelud trava pa su tako trave zbog velikog broja vrsta i široke rasprostranjenosti jedan od glavnih izvora peludi koje uzrokuju alergije. Razdoblje cvatnje trave traje od travnja do rujna te je tako dulje od cvatnje ostalih biljaka koje uzrokuju alergije. Vrhunac cvata trave događa se u svibnju i tada su koncentracije peludi najviše. Na slikama 11., 12. i 13. prikazane su predikcije(plava boja) i stvarne vrijednosti(narančasta boja) breze za godine 2019., 2020. i 2021. Na slikama je vidimo da su koncentracije najviše za vrijeme početka cvata te kasnije blago opadaju. Kako bismo bolje shvatili problem predikcije peludi trave, navest ćemo neke od najčešćih trava koje izazivaju alergije te njihovo predviđeno vrijeme cvatnje([10]):

- Klupčasta oštrica (lat. *Dactylis glomerata*) – cvate od svibnja do kolovoza
- Medunika (lat. *Holcus sp.*) – cvate u lipnju i srpnju
- Livadna mačica (lat. *Phleum pratense*) – cvate od svibnja do rujna
- Ljuljevi (lat. *Lolium perenne*) – cvatu od svibnja do rujna
- Vlasulje (lat. *Festuca sp.*) – cvatu od ožujka do listopada
- Obična zubača (lat. *Cynodon dactylon*) – cvate od lipnja do kolovoza

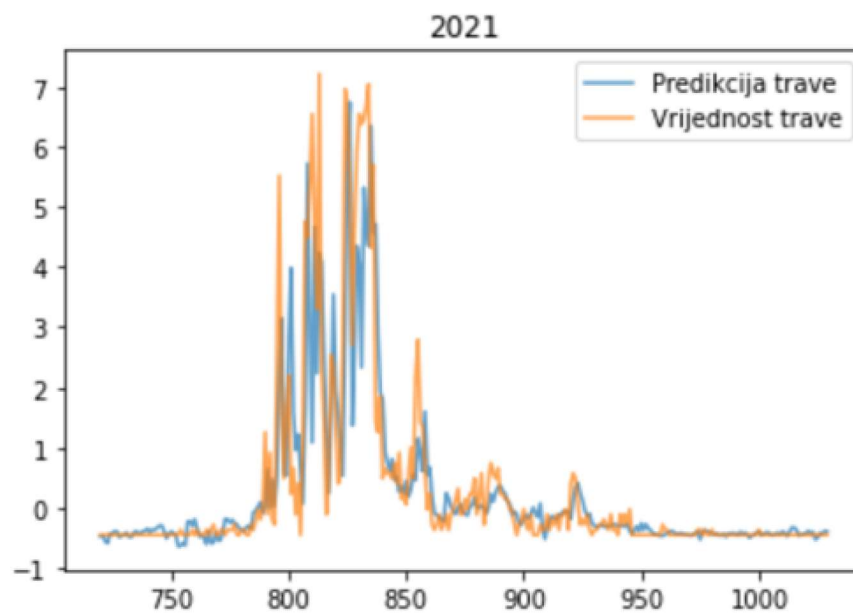
Stoga, zbog praćenja velikog broja vrsta, točne koncentracije peludi trava u zraku teško su predvidive, no model i dalje može dati dobru predodžbu o tome kako će se koncentracije ponašati.



Slika 11: Predikcije i stvarne koncentracije peludi trave u 2019.godini



Slika 12: Predikcije i stvarne koncentracije peludi trave u 2020.godini

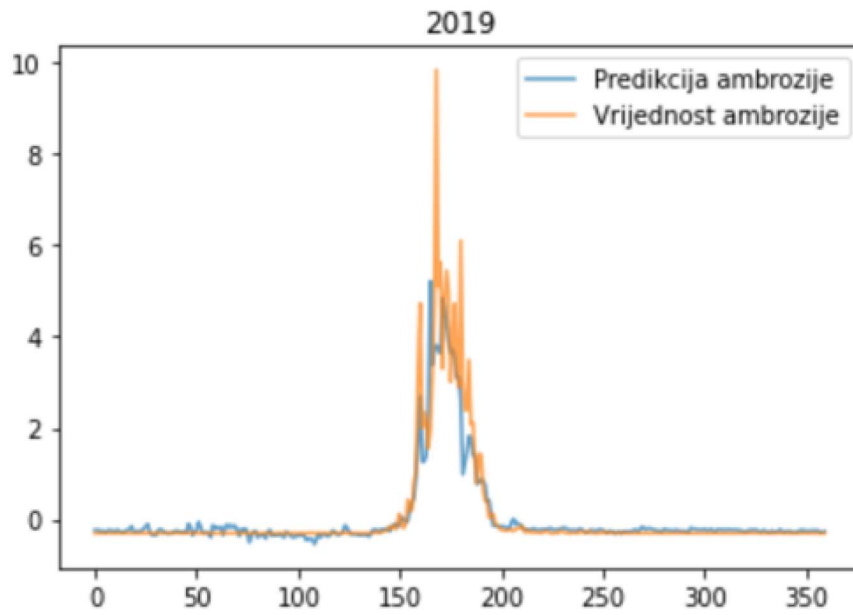


Slika 13: Predikcije i stvarne koncentracije peludi trave u 2021.godini

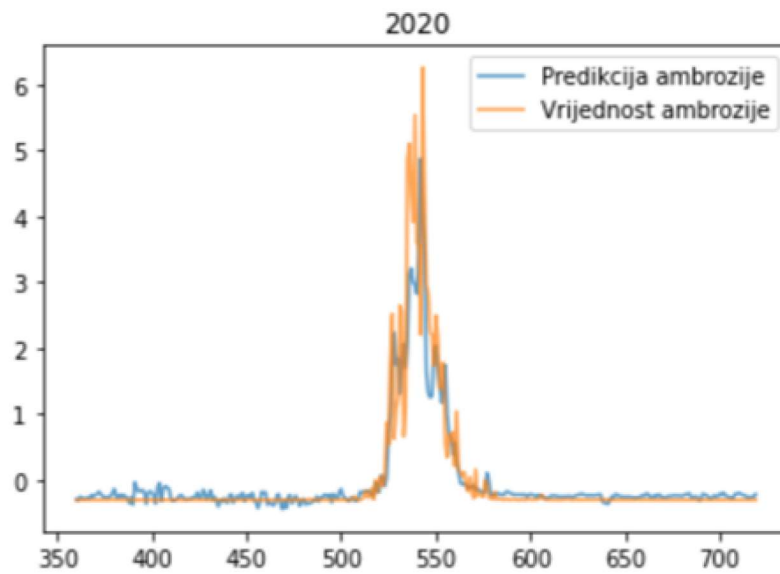
### 5.3 Pelud ambrozije

Ambrozija je biljka koja raste na zapuštenim zemljištima i livadama, uz putove, polja te među usjevima i u vrtovima. Jedna je od biljaka koje proizvode najveću količinu peludi. Dok ostale biljke uzrokuju alergijske reakcije pri vrlo visokim koncentracijama, ambrozije ih uzrokuje već pri svojoj pojavi. Ambrozija počinje cvjetati tijekom srpnja, a vrhunac doseže tijekom kolovoza i početka rujna, nakon čega cvat prestaje te joj koncentracija opada. ([11]) Na slikama 14., 15. i 16. prikazane su predikcije(plava boja) i stvarne vrijednosti(narančasta boja) ambrozije za godine 2019., 2020. i 2021. Vidimo da koncentracije naglo rastu, dosežu vrhunac te zatim naglo opadaju. Baš kao kod peludi breze, model radi najveće greške kada su stvarne vrijednosti vrlo visoke, no predviđene vrijednost ipak približno rastu i padaju kao

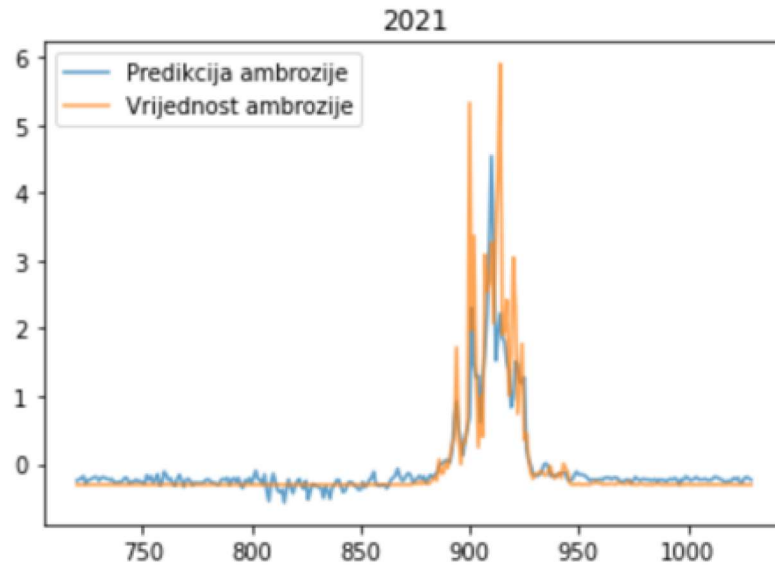
i stvarne te stoga mogu poslužiti kao pomoć ljudima koji imaju alergiju na ambroziju.



Slika 14: Predikcije i stvarne koncentracije peludi ambrozije u 2019.godini

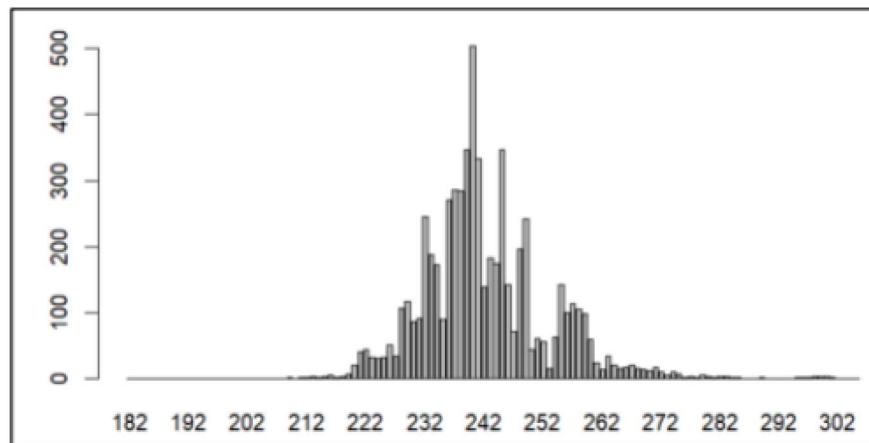


Slika 15: Predikcije i stvarne koncentracije peludi ambrozije u 2020.godini

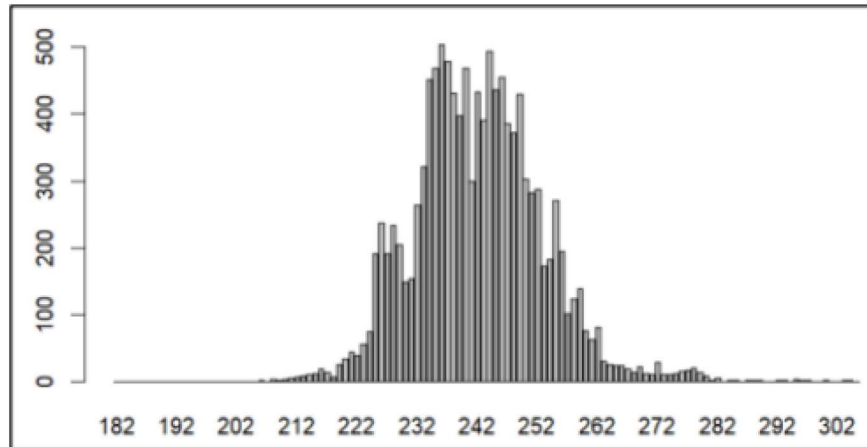


Slika 16: Predikcije i stvarne koncentracije peludi ambrozije u 2021.godini

Također treba spomenuti još dva grafa rađena sa stvarnim vrijednostima ambrozije (slike 17. i 18.).



Slika 17: Prosjek koncentracije ambrozije po danu za godine 2010., 2011. i 2012.



Slika 18: Prosjek koncentracije ambrozije po danu za godine 2019., 2020 i 2021.

Na stupičastim dijagramima prikazana je koncentracija ambrozije po rednom broju dana. Dani su uzeti iz srpnja, kolovoza, rujna i listopada, što su mjeseci kada je ambrozija prisutna u zraku. Na slici 17. prikazan je prosjek koncentracije ambrozije po danu za godine 2010., 2011. i 2012. (nismo htjeli uzeti točno određenu godinu već smo, kako bismo što bolje i točnije dočarali približne vrijednosti za to razdoblje, za svaki dan na grafu prikazali aritmetičku sredinu vrijednosti toga dana za tri uzastopne godine). Prosječna koncentracija ambrozije za to razdoblje iznosi 49.5014. Slika 18. prikazuje podatke na isti način, ali za godine 2019., 2020 i 2021. Prosječna koncentracija ambrozije za to razdoblje iznosi 96.2846, što je gotovo duplo više nego u razdoblju prije 10 godina, a što je također moglo utjecati na točnost modela.

## 6 Zaključak

Peludne alergije problem su koji muči veliki broj ljudi, a sve većim napretkom znanosti i strojnoga učenja sigurno će se moći naći model koji će vrlo točno predviđati koncentracije peludi u zraku te tako pomoći ljudima koji pate od alergije.

Osim mogućeg pronalaska boljeg modela za rješavanje ovoga problema, osvrnimo se na to da su pogreške u predikcijama moguće i zbog velikih klimatskih promjena posljednjih godina koje direktno utječu na koncentraciju peludi u zraku (vidi slike 17. i 18.).

Iako ne možemo uvijek predvidjeti u potpunosti točne rezultate, model će svejedno predvidjeti kada su koncentracije peludi u zraku više od graničnih vrijednosti koje izazivaju alergiju te tako može pomoći osobama koje pate od alergija upozoravajući ih na visoke koncentracije.

## Literatura

- [1] A. Mittal, *Understanding RNN and LSTM*, Medium, Web izvor dostupan na: <https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e>
- [2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, MIT Press, 2016. <http://www.deeplearningbook.org>
- [3] Yi-Ting Tsai, Yu-Ren Zeng, and Yue-Shan Chang, *Air Pollution Forecasting Using RNN with LSTM*, In: *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, 2018, pp. 1074-1079. DOI: 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00178
- [4] <https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/cor.test> (19.rujan 2023.)
- [5] <https://ianlondon.github.io/blog/encoding-cyclical-features-24hour-time/> (19.rujan 2023.)
- [6] <https://proclusacademy.com/blog/explainer/regression-metrics-you-must-know/> (19.rujan 2023.)
- [7] <https://optimization.cbe.cornell.edu/index.php?title=Adam> (19.rujan 2023.)
- [8] TensorFlow dokumentacija: [https://www.tensorflow.org/api\\_docs](https://www.tensorflow.org/api_docs) (19.rujan 2023.)
- [9] [https://www.zzjz-kkz.hr/default\\_vijest.asp?sid=7975](https://www.zzjz-kkz.hr/default_vijest.asp?sid=7975) (19.rujan 2023.)
- [10] <https://stampar.hr/hr/novosti/visoko-alergena-pelud-trava> (19.rujan 2023.)
- [11] <https://www.plivazdravlje.hr/alergije/alergeni/13/Ambrozija.html> (19.rujan 2023.)
- [12] Git repozitorij : <https://github.com/zmarinko/ZavrsniRad>