

ПРЕДВИЂАЊЕ БРОЈА СКИЈАШКИХ ПОВРЕДА КОРИШЋЕЊЕМ
АЛГОРИТАМА ВРЕМЕНСКИХ СЕРИЈА И МАШИНСКОГ УЧЕЊА
COMPARISON OF TIME SERIES AND MACHINE LEARNING
ALGORITHMS FOR PREDICTING SKI INJURIES

Јана Радивојевић, Сандро Радовановић, Милица Булајић, Борис Делибашић

РЕЗИМЕ: Скијање је јако популаран спорт, а због могућности развоја велике брзине и све чешћег извођења акробатских елемената од стране скијаша постоји ризик од озбиљних повреда. У циљу унапређења процеса доношења одлука, вођење евиденције и предвиђање скијашких повреда је постало веома значајно за скијалишта. У раду је илустрована примена и поређење перформанси модела временских серија као и модела машинског учења у предвиђању броја скијашких повреда на Копаонику. Предвиђан је број повреда на дневном, недељном и месечном нивоу. Предвиђање повреда на различитим временским фреквенцијама пружа бољи увид у целокупно кретање броја повреда током времена, што свакако утиче на квалитет планирања и доношење одлука. Целокупно истраживање је спроведено у програмском језику Пајтон.

КЉУЧНЕ РЕЧИ: скијашке повреде, предиктивни модели, временске серије, Копаоник.

ABSTRACT: Skiing is a very popular sport, due to the possibility of high-speed development and frequent performance of acrobatic elements by skiers, there is a risk of serious injuries. To improve the decision-making process, keeping records and predicting ski injuries has become very important for ski resorts. This paper illustrates the application and comparison of the time series model and classical machine learning models in predicting the number of ski injuries on Kopaonik. The number of injuries is predicted on a daily, weekly and monthly basis. Predicting injuries at different time frequencies provides a better insight into the overall movement of the number of injuries over time, which certainly affects the quality of planning and decision making. The entire research was conducted in the Python programming language.

KEY WORDS: Key words: ski injuries, predictive models, time series, Kopaonik.

УВОД

Скијање је један од најпопуларнијих зимских спортова. У земљама које имају адекватне услове управо су скијалишта главни извор прихода у планинским регионима. У скијашкој индустрији у Србији водећу улогу је преузео Копаоник, са надморском висином од 2017 метара, више од 50 километара скијашких стаза, са 23 жичаре, капацитетом од 32000 скијаша на сат и око 160 снежних дана у години. Скијање спада међу високо ризичне спортове због потенцијалне опасности од тешких повреда. Како наводе аутори у раду [10], од 1970 године до данас број скијашких повреда опао је са 5-8 на 2-3 повреде на 1000 скијашких дана. Дакле, повреда скијаша се може сматрати ретким догађајем. Ипак, оне које се десе могу имати велике, а некад и фаталне последице по скијаша.

Из тог разлога побољшање учинка и управљање ризиком од повреда постао је главни циљ свих већих скијалишта, па тако и Копаоника. За управу скијалишта је од значаја да предвиди број повреда како би се адекватно управљало ресурсима у смислу обезбеђивања довољно спасилачког особља и опреме и како би се спровеле адекватне превентивне акције. Са друге стране, непружање адекватне помоћи скијашима могло би довести до нарушавања репутације скијалишта и губитка поверења. Истраживања у области скијашких повреда претежно се баве идентификовањем фактора који повећавају ризик од настанка повреде. У раду [13] дата је анализа ових фактора на основу извештаја Горске службе о повредама на Копаонику током сезоне 2016/2017 и утврђена је повезаност врсте повреде, локације, употребе заштитне опреме и метеоро-

лошких услова. Дошло се до занимљивих закључака да се највећи број скијаша повредио на добро сређеним стазама, у сунчаним интервалима и у другој половини скијашког дана, после 13 часова. Аутори у раду [6] дају поделу фактора на основу тога да ли зависе од скијалишта или не. У прву групу, фактора који не зависе од скијалишта, спадају личне карактеристике скијаша, њихово здравствено стање, старост, брзина којом скијају, умор, квалитет опреме, скијашко искуство, итд. У групу фактора које зависе од конкретног скијалишта су припрема и квалитет стаза, снежни услови, кумулативна скијашка статистика као и временски услови. Ови фактори су од посебног значаја за унапређење стратегије скијалишта у смањењу ризика од повреда. Такође, скијалишта могу да обезбеде податке о улазима на жичару, који могу бити од значаја у предвиђању скијашких повреда. Ови подаци су од значаја јер дају информације о кретању скијаша. Из овог разлога се у све већем броју истраживања ови подаци додају осталим подацима о повредама. Аутори [2] су на основу ових података предложили модел за превиђање ризика од повреда у скијалиштима заснован на искуству доносиоца одлуке. Модел је направљен као систем за подршку одлучивању који користити хијерархијску организацију особина која описује ризик повређивања на дневном нивоу на скијалишту. Тачније, развијен је *DEX* модел који се користи за описивање ризика повређивања на скијалишту на дневном нивоу. Овај модел је додатно унапређен у раду [7]. Битан закључак овог рада је да је недостатак модела машинског учења да препознају структуре и односе између атрибута могуће надоместити интеграцијом доменског знања стручњака. Визуализацијом података могуће је упоредити

понашање скијаша које доводи до повреде и оно које не доводи. У раду [15] аутори предлажу просторно-временски модел за предикцију броја повреда на скијалишту на основу података о транспорту и метеоролошких података. Овај модел омогућава нешто раније предвиђање повреде, што је значајно због преузимања адекватне акције. У раду [5] на основу података о транспорту скијаша жичарама и метеоролошким условима изграђено је неколико модела машинског учења у циљу предвиђања скијашких повреда. Најбоље резултате дао је модел заснован на алгоритму градијентом подстакнутих стабала (енгл. *Gradient Boosted Trees*). Аутор ово објашњава постојањем нелинеарних односа у подацима, те да из тог разлога модел логистичке регресије није био адекватан избор. Такође, истиче се да је могуће да предиктивни модели не дају задовољавајуће резултате због постојања дисбаланса класа. Како је повреда скијаша редак догађај, преузорковање (енгл. *oversampling*) би могло да побољша перформансе ових модела. Што се тиче будућих трендова истраживања у овој области, аутори горе наведених радова сматрају да је потребно усредсредити се на развој предиктивних, персонализованих и превентивних модела. Овакав оквир примењен у пракси био би додатна подршка у доношењу одлука у скијалиштима.

У овом раду се овом проблему приступа на специфичан начин. Историјски подаци о повредама скијаша на Копаонику моделовани су као временске серије. Дакле, предвиђање броја повреда у будућности засниваће се искључиво на прошлим вредностима. Овим се свесно занемарују сви остали фактори, што свакако представља главно ограничење оваквог приступа. Циљ је испитати какви се резултати могу добити на овакав начин, као и упоредити резултате модела временских серија и модела машинског учења, у циљу проналажења оног који даје најпрецизније резултате. У наставку рада биће наведени основни теоријски концепти временских серија и њиховог моделовања, након тога укратко ће бити описан поступак изградње предиктивних модела као и њихова евалуација. На самом крају биће упоређени добијени резултати.

МЕТОДОЛОГИЈА ИСТРАЖИВАЊА

Временске серије

Временска серија представља уређени низ опсервација. Уређивање се најчешће врши у односу на време и то у једнаким временским интервалима [14]. Оно што издваја анализу временских серија као статистичку дисциплину јесте то што се временска серија посматра као скуп случајних променљивих за које се претпоставља да су међусобно зависне и веома често и корелисане. Дакле, временске серије у модел уводе још једну, временску, димензију. Управо због постојања временске међузависности могуће је формирати моделе временских серија који прогнозирају будуће вредности само на основу прошлих. Временске серије имају јако широку примену у пракси, у областима

које укључују мерење неке појаве кроз време, као што су економија, метеорологија, финансије, итд. Временске серије су моћан алат јер омогућавају анализу прошлости, праћење садашњости и предвиђање будућности. Међутим, у раду са временским серијама треба бити опрезан како њихова специфична својства обично траже специјализоване статистичке методе и моделе. [12]

Класификацију временских серија је могуће урадити на основу различитих критеријума. Кључна је подела на стационарне и нестационарне временске серије. Ова подела прави разлику у понашању временске серије у току времена. Уколико се временска серија понаша по устаљеном обрасцу током времена она је стационарна. Супротно, уколико је понашање временске серије дато као функција времена, ради се о нестационарној временској серији [11].

Свака временска серија се састоји из компоненти, а на основу релација између њих изводе се различите претпоставке и гради модел који најбоље описује временску серију. Компоненте временске серије су тренд, сезоналност, цикличност и случајна компонента. Тренд је важна компонента која представља дугорочну тенденцију развоја временске серије. Он описује понашање временске серије за њен читав временски период. У зависности од тога да ли серија опада или расте кроз време, разликујемо опадајући и растући тренд. Још једна подела је на детерминистички и стохастички тренд, у односу на то да ли се кретање временске серије може предвидети на основу претходних вредности. Детерминистички тренд је могуће описати математичком функцијом. Постоје различити модели тренда у односу на избор функције: линеарни, параболички, експоненцијални, логистички итд. Уколико временска серија има стохастички тренд он се не може описати и њено понашање се самим тим не може предвидети. [12] Поред тренда, као детерминистичка компонента, може се јавити и сезоналност. Сезоналност представља правилно одступање од средње вредности серије које се редовно понавља кроз временски период. На пример, годишње или месечно [8]. Временска серија садржи и случајну компоненту која се назива бели шум. Он представља насумичну промену у серији која се не може образложити и у његовом настајању нема правилности. Ова компонента се понекад назива у резидуална компонента. Веза између компоненти временске серије може бити адитивна, мултипликативна или комбинација адитивне и мултипликативне. Дакле, модели временских серија се добијају множењем или сабирањем горе описаних компоненти. [3]

Главни циљ моделовања временских серија је пажљиво прикупљање и студирање прошлих опсервација како би се изградио одговарајући модел који најбоље одговара структури временске серије и самим тим најбоље предвиђа њене будуће вредности [1]. Постоје различити модели временских серија, а то који ће дати најбоље резултате при предвиђању зависи од конкретног проблема и података који се користе за његово решавање, па је пре одабира модела пожељно урадити анализу временске серије. Треба истаћи разлику између анализе и прогнозе временске серије, иако

се ови појмови често поистовећују. Анализа је усмерена на саму структуру временске серије и њене карактеристике. Анализа временске серије нам даје информације о њеном понашању, што подразумева евентуални раст или опадање вредности, циклично понашање, као и велика одступања у вредностима. Са друге стране, прогноза временских серија је употреба одређеног модела за предвиђање будућих вредности на основу прошлих вредности, а то се може дефинисати и као метода предвиђања будућих вредности разумевањем прошлих вредности. [4]

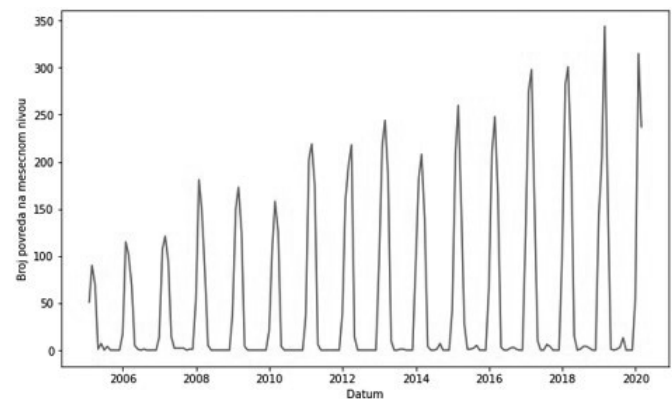
Након детаљне анализе понашања временске серије могуће је изабрати оптималан модел за предвиђање будућих вредности, а како је данас познат велики број модела временских серија, избор је велики и треба му посветити доста пажње јер од њега зависи успех предвиђања будућих вредности [9]. Неки од најчешће коришћених модела у раду са временским серијама су: ауторегресивни модел (енгл. *Autoregression, AR*), модел покретних средина (енгл. *Moving average, MA*), интегрисани ауторегресивни модел покретних средина (енгл. *Autoregressive Integrated Moving Average Model, ARIMA*) и његова модификација сезонски интегрисани ауторегресивни модел покретних средина (енгл. *Seasonal Autoregressive Integrated Moving Average Model, SARIMA*). Треба напоменути да је код свих модела временских серија специфична подела података на оне за учење и тестирање модела. Као што је раније назначено, у временској серији опсервације нису независне, па не можемо податке расподелити насумично као што то иначе чинимо. Карактеристике података временских серија, као што су ауторегресивна природа, тренд или сезоналност онемогућавају случајну поделу. На пример, ако су наше опсервације аутокорелисане, то би био проблем зато што би модел који учи над прошлом, нормално добро предвиђао следећу аутокорелисану опсервацију. У том случају, добијени резултати би били превише оптимистични у погледу предиктивне снаге модела, тако да се подела углавном врши у повезаним нивовима података. Након одабира модела и поделе података, битно је дефинисати хиперпараметре модела. Ово је јако битан корак како у многоме одређује квалитет модела, како ће модел прогнозирати нове и невиђене примере тј. колика му је моћ генерализације. [3] Након изградње модела и предвиђања будућих вредности, врло је битно одредити њихову тачност. Најчешће коришћене мере тачности предвиђања временских серија су просечно апсолутно растојање (енгл. *Mean absolute error, MAE*), корен средњег квадратног одступања (енгл. *Root mean squared error, RMSE*) и просечни проценат апсолутног одступања (енгл. *Mean absolute percentage error, MAPE*). [1]

Поставка истраживања

Пре самог одабира и изградње модела временских серија, извршена је припрема података, анализа података, као и анализа компоненти временске серије. Сва пред-

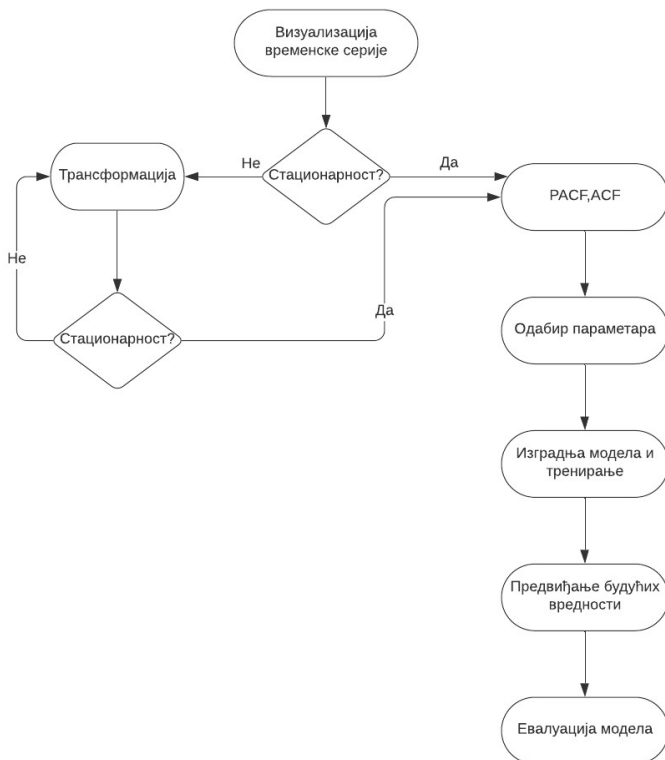
виђања у раду вршена су на основу историјских података о повредама на Копаонику у периоду од јануара 2005 закључно са фебруаром 2020 године. Скуп података садржи историјску евиденцију броја повреда за временски период од 5.529 дана тј. 182 месеца тј. 791 недељу.

Временске серије имају специфичну структуру, за моделовање временске серије потребна је само временска променљива и променљива коју предвиђамо. Како је потребно назначити која променљива представља временску димензију, неходно је поставити временски индекс (енгл. *DatetimeIndex*). Индексирање омогућава лако филтрирање података на годишњем, месечном, дневном, на нивоу недеље, према одређеним данима итд. Пожељно је визуализовати временску серију, да би се уочило њено кретање током времена, као и изражене карактеристике. Графички приказ временске серије која је овде анализирана, на месечном нивоу, дат је на слици 1.



Слика 1. Временска серија

Као што је раније наведено, стационарност је важна карактеристика временске серије, па је на почетку неопходно испитати стационарност серије која се овде користи. Иако се преко графичког приказ серије и кретање стандардне девијације и покретних средина током времена може наслутити да ли се ова својства мењају током времена, тј. да ли је серија стационарна, закључак је боље донети помоћу Дики-Фулера теста (енгл. *Dickey-Fuller Test*). Нулта хипотеза овога теста је да је серија нестационарна. Изненађујуће, овде анализирана временска серија је, према резултатима Дики-Фулера теста ($p < 0.05$) стационарна на дневном и недељном нивоу и поред изразите компоненте сезоналности и постојања тренда који се уочавају на графичком приказу података. Међутим, на месечном нивоу то није случај како p вредност износи 0,3970 закључује се да серија није стационарна и потребно је извршити одређене трансформације како би је довели у стационарно стање. То је извршено диференцирањем. Са обзиром да имамо изразиту сезоналност у подацима, потребно је користити сезонски интегрисани ауторегресивни модел покретних средина, SARIMA модел (енгл. *Seasonal Autoregressive Integrated Moving Average Model, SARIMA*). Основни кораци који су спорведени у изградњи овог модела представљени су на слици 2.



Слика 2. Основни кораци изградње SARIMA модела

SARIMA модел се записује као $ARIMA(p, d, q)(P, D, Q, m)$. Где су параметри p, d, q несезонски, а параметри P, D, Q њима еквиваленти на сезонском нивоу и параметар m који означава трајање сезоне. Параметар p означава број временски померених вредности које ће бити укључене у модел, q је број временски померених грешака предвиђања у функцији предвиђања и параметар d означава ред несезонског диференцирања. У одређивању ових параметара може нам помоћи тумачење графикана аутокорељационе (енгл. Autocorrelation function, ACF) и парцијалне аутокорељационе функције (енгл. Partial autocorrelation function, PACF) временске серије оног нивоа диференцирања на којем је услов стационарности испуњен. Можемо израчунати корелацију за посматрање опсервације временске серије са опсервацијама у претходним временским периодима, званим временски померене вредности (енгл. lagged values). Будући да се корелација посматраних временских серија израчунава са вредностима исте серије у претходним временским тренуцима, она се назива серијска корелација или аутокорељација. Парцијална аутокорељација поред овога посматра и корелације са временских померених вредностима које нису узете у обзир претходним временски помереним опсервацијама. На основу графикана аутокорељационе функције одређује се параметар q , док се за процену параметра p користи график парцијалне аутокорељационе функције [3]. Посебну пажњу треба обратити на вредности које се налазе ван интервала поверења, како оне одређују потенцијалне вредности параметра. Ово би било довољно за изградњу ARIMA модела. За изградњу SARIMA модела, што је овде циљ, недостају сезонски параметри модела. Ови параметри се не могу најпрецизније оценити на основу PACF и ACF за

неколико сезонских интервала. Из овог разлога, за оптимизацију сезонских параметара модела коришћена је мрежна претрага (енгл. GridSearch). Пре саме имплементације претраге било је неопходно дефинисати трајање сезоне на свим фреквенцијама, тј. параметар m . Ове вредности добијене су анализом графикана аутокорељације за различити број временски померених вредности како би се утврдило за који период се понашање серије понавља. Уколико је сезоналност изражена на графикону аутокорељације ће бити изражене вредности у тим временским периодима, на пример код месечних предикција где m износи 12 оне ће се налазити на временским вредностима од 12, 24, 36 итд. За недељни ниво утврђено је да m износи 52, док на дневном нивоу m износи 70. Такође, пре претраге је потребно одредити хоризонт предвиђања и на основу тога извршити поделу података на тренинг и тест скуп. Одабрани период за који ће се предвиђати број повреда биће издвојен као тест скуп, док ће остатак да се користи за учење модела. Предвиђан је број повреда за временски хоризонт од 166 дана, 4 недеље и 2 месеца. Мрежна претрага итеративно проверава различите комбинације параметара. Потребно је задати интервал у оквиру ког ће се вршити претрага параметара, у његовом одређивању може нам помоћи графикон аутокорељације. За сваку комбинацију гради се и тренира нови модел функцијом $SARIMAX()$ из библиотеке $statsmodels$ и процењује се његов квалитет. Када се пробају све комбинације параметара, као оптимална комбинација изабраће се она која даје најбоље резултате тј. најбоље задовољава критеријум од интереса. У нашем случају, то је Акаиков информациони критеријум (енгл. Akaike's Information Criterion, AIC), који треба бити минималан. Њега је лако издвојити како је већ уграђен у модел и он мери колико модел одговара подацима, узимајући у обзир сложеност модела. Од интереса је пронаћи модел који даје најбоље резултате у поређењу са његовом сложености. Након извршене претраге пронађени су оптимални параметри за моделе, њихов приказ дат је на слици 3. За предвиђање је коришћена метода $forecast()$, којој је као параметар потребно проследити број корака за које се врши предвиђање.

	Дневни ниво	Недељни ниво	Месечни ниво
p	1	1	1
d	1	1	1
q	1	1	1
P	0	1	0
D	1	1	1
Q	1	1	1
m	70	52	12

Слика 3. Параметри модела

На основу горе приказаног поступка изградње модела временских серија може се доћи до закључка да припрема података, анализа временске серије, одабир модела а посебно одабир одговарајућих параметара могу одузети доста времена, тако да је у већини случаја лакше одабрати неколико карактеристика из постојеће временске серије и њих моделовати као моделе машинског учења. У сврху упо-

ређивања добијених предикција ова два приступа, изграђени су модели линеарне регресије (енгл. *Linear regression*) и случајних шума (енгл. *Random forest*) за временске серије на свим раније анализираним фреквенцијама. Дакле, укупно шест модела а поступак њихове изградње и тренирања биће илустрован у наставку рада.

Први корак јесте додавање атрибута постојећим временским серијама. Начин на који могу да се извуку додатне информације из постојеће серије је посматрање временских померених вредности, као и односа временских померених вредности и њихове статистике. Овде је битно одредити који временски оквир тј. прозор (енгл. *window*) временских померених вредности је потребно укључити у анализу. Овде не постоји правило за одабир, већ се до њега долази применом различитог броја корака у рачунању временских померених вредности. За дневне предикције коришћен је оквир од **5 корака** уназад, за недељне **8 корака** и за месечне оквир од **2 корака** уназад. Након додавања временски померених вредности као атрибуту додати су и њихова средња, максимална и минимална вредност. Међутим, додавање ових атрибута позитивно је утицало на квалитет предвиђања код дневних и недељних модела, док код модела са месечном фреквенцијом то није био случај, те су они уклоњени из истог. У раду са временски помереним вредностима временске серије треба водити рачуна о непознатим вредностима које се јављају за првих k опсервација, где је k број корака уназад при рачунању временски померених вредности. Како модели не могу да раде са непознатим вредностима, неходно је уклонити ове опсервације. Након додавања атрибута и сређивања података, врши се њихова подела на скуп за тренирање и учење. Затим следи изградња и тренирање модела. За рачунање предвиђања оба модела коришћена је метода *predict()*, којој се као параметар прослеђује скуп података за проверу. Треба напоменути да је за предвиђања ових модела коришћен исти хоризонт предвиђања коришћен код SARIMA модела, ради лакшег поређења добијених резултата.

РЕЗУЛТАТИ И ДИСКУСИЈА

Као критеријум за оцењивање квалитета предикције свих девет изграђених модела коришћен је корен средње квадратне грешке (енг. *Root mean squared error, RMSE*). Ова грешка се рачуна применом функције *sqrt()* на средњој квадратној грешци добијеној применом методе *mean_squared_error()*. Добијени резултати табеларно су приказани на слици 4.

RMSE	Дневне предикције	Недељне предикције	Месечне предикције
SARIMA	6.507	13.775	91.662
Линеарна регресија	2.486	15.532	177.468
Случајне шуме	2.575	17.314	82.925

Слика 4. RMSE грешке

Иако би се на први поглед на основу добијених резултата могло рећи да је дневни модел најпрецизнији, тај закључак не би био исправан како не можемо занемарити чињеницу да се ради о моделима са различитим временским фреквенцијама као и различитим хоризонтима предвиђања. Да би резултати горе наведених модела били међусобно упоредиви потребно је задати им исти временски хоризонт предвиђања. На тај начин би се открило на којој фреквенцији предвиђања се добијају најбољи резултати тј. најтачнија предвиђања броја скијашких повреда. Ипак, то није циљ овог рада, већ поређење перформанси SARIMA модела и модела машинског учења.

За дневне предикције најмања RMSE грешка добијена је применом модела линеарне регресије док мало већу грешку прави модел случајних шума. Међутим, у поређењу са SARIMA моделом, њихова грешка је мања за око 4 на дневном нивоу, што представља значајну разлику. Будући да нам се стварне вредности за задати хоризонт предвиђања крећу од 0 до 23, RMSE вредност од 2.5 је задовољавајућа. Дакле за дневна предвиђања броја повреда не препоручује се примена SARIMA модела.

Другачија ситуација је са моделима за недељна предвиђања, како управо SARIMA модел даје најбоље предикције. Код недељних предикција класични алгоритми нису дали очекивано добре резултате, чак ни са оптимизацијом параметара није добијена мања грешка од оне коју прави сезонски интегрисани ауторегресивни модел покретних средина. На побољшање перформанси модела вероватно би утицало повећање временског оквира заоставних вредности које посматрамо (који је износио 8 недеља у овом случају).

Код месечних предвиђања постоји велика разлика у вредности грешке модела линеарне регресије и случајних шума. Алгоритам случајних шума дао је боље вредности и од SARIMA модела, тако да он свакако представља најбољи избор код ових предвиђања.

Међутим, поред саме грешке модела у обзир се при оцењивању извршеног моделовања требају узети и остале карактеристике. На пример, са аспекта брзине моделовања моделима машинског учења треба дати предност, како према података и оптимизација параметара при изградњи SARIMA модела одузима доста времена. Такође, како би моделовање SARIMA моделом било могуће, подаци морају задовољити полазне претпоставке о линеарности и стационарности. Са друге стране уколико нас посебно занима структура временске серије, њено понашање и из самог модела желимо добити што више информација, бољи избор су модели временских серија како су они специјализовани за рад са овако структурираним подацима.

Моделовање временских серија моделима машинског учења показало се као доста једноставније. Највећи изазов је представљао одабир одговарајућег оквира временски померених вредности, како су ручно пробане различите вредности. Ипак треба нагласити да резултати модела машинског учења зависе од количине података са којима се располаже. Како се са продужавањем интервала на који се подаци односе смањује количина података, код резултата добијених на недељном и месечном нивоу, наслућује се евентуална претренираност модела.

ЗАКЉУЧАК

Кроз рад је описан поступак изградње и тренирања модела специјализованих за рад са временским серијама али и модела машинског учења. Поступак изградње SARIMA модела је неупоредиво дужи и компликованији у односу на моделе машинског учења. Сам процес оптимизације параметара је најбитнији, али и најдуготрајнији. Мрежна претрага параметара је јако спора, а одабир одговарајућег опсега претраге је тежак, поготово сезонских параметара када на располагању имамо историјске податке за дужи временски период. Вероватно ни сам одабир опсега у овом раду није оптималан, и нека друга комбинација параметара би евентуално дала прецизнија предвиђања. Треба напоменути да су у рад укључене оптималне комбинације параметара од свих пробаних на подскупу података за учење модела, што не значи да су оне најбољи могући избор. Приступ примене модела машинског учења дао је боље резултате у већини случајева, осим код недељних предикција, где је добијена нешто већа вредност RMSE грешке. Уколико се упореди време потребно за изградњу и учење модела са квалитетом тј. прецизношћу добијених предикција, предност свакако треба дати моделима машинског учења. Оно што је недостатак учења временских серија моделима машинског учења је отежана интерпретације резултата, како су додати атрибути временске серије нетипични за ове моделе.

Можемо сматрати да је циљ истраживања испуњен како су добијене дневне, недељне и месечне предикције броја скијашких повреда за жељени временски период. Занимљиво је да су за све три временске фреквенције најбоља предвиђања добијена применом различитих алгоритама, ово још једном потврђује значај посматрања проблема предвиђања скијашких повреда са различитих временских димензија. За дневне предикције најбоље је користити модел линеарне регресије, за недељне сезонски интегрисани ауторегресивни модел покретних средина, док је за месечне предикције најбољи модел случајних шума.

ЛИТЕРАТУРА

- [1] Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613.
- [2] Bohanec, M., & Delibašić, B. (2015, May). Data-mining and expert models for predicting injury risk in ski resorts. In International conference on decision support system technology (pp. 46-60). Springer, Cham.
- [3] Brownlee, J. (2017). Introduction to Time Series Forecasting With Python: How to Prepare Data and Develop Models to Predict the Future. Machine Learning Mastery
- [4] Chen, P., Niu, A., Liu, D., Jiang, W., & Ma, B. (2018, August). Time series forecasting of temperatures using SARIMA: An example from Nanjing. In IOP Conf. Ser. Mater. Sci. Eng (Vol. 394, No. 5, pp. 1-7).
- [5] Dallagiaco, M. (2017) Predicting the risk of accidents for downhill skiers. Master of Science Thesis, School of Information and Communication Technology, KTH Royal Institute of Technology, Stockholm, Sweden
- [6] Delibašić, B., Makajić-Nikolić, D., Ćirović, M., Petrović, N., & Suknović, M. (2020). A ski injury risk assessment model for ski resorts. Journal of Risk Research, 1-13.
- [7] Delibašić, B., Radovanović, S., Jovanović, M., Bohanec, M., & Suknović, M. (2018). Integrating knowledge from DEX hierarchies into a logistic regression stacking model for predicting ski injuries. Journal of Decision Systems, 27(sup1), 201-208.
- [8] Ivčec, A. (2018). Prognoza vremenskih serija u senzorskim tokovima podataka. Diplomski rad, Univerzitet u Zagrebu, Fakultet elektronike i računarstva.
- [9] Knežević, T. (2013). Analiza fazi vremenskih serija. Master rad, Univerzitet u Novom Sadu, Prirodno-matematički fakultet.
- [10] Koehle, M. S., Lloyd-Smith, R., & Taunton, J. E. (2002). Alpine ski injuries and their prevention. Sports Medicine, 32(12), 785-793.
- [11] Kovačić, Z. (1995). Analiza vremenskih serija. Ekonomski fakultet, Beograd.
- [12] Mihailović, A. (2011). Analiza vremenskih serija. Master rad, Univerzitet u Novom Sadu, Prirodno-matematički fakultet.
- [13] Minčić, S., Stojiljković, N., Pavlović, L., Pantelić, S., Sporiš, G., & Krističević, T. (2018, March). Analysis of skiing and snowboarding injuries-science in the service of injuries prevention. In 2nd Scientific Conference SPE BALKAN SKI: Science, Practise & Education.
- [14] Momirović, M. (2018). Specifičnost vremenskih serija u ekonomiji. Master rad, Univerzitet u Beogradu, Matematički fakultet.
- [15] Radovanovic, S., Delibasic, B., Suknovic, M., & Matovic, D. (2019). Where will the next ski injury occur? A system for visual and predictive analytics of ski injuries. Operational Research, 19(4), 973-992.



Јана Радивојевић, Универзитет у Београду – Факултет организационих наука

Контакт: janaradivojevic5@gmail.com
Област интересовања: Пословна интелигенција, машинско учење, пословно одлучивање, складиштење података



Сандро Радовановић, Универзитет у Београду – Факултет организационих наука

Контакт: sandro.radovanovic@fon.bg.ac.rs
Област интересовања: Машинско учење, Откривање законитости у подацима, Складишта података, Системи за подршку одлучивању, Пословна интелигенција, Теорија одлучивања



Милица Булајић, Универзитет у Београду – Факултет организационих наука

Контакт: milica.bulajic@fon.bg.ac.rs
Област интересовања: Теорија вероватноће, Статистика, Мултиваријациона анализа, Економетрија



др Борис Делибасић, редовни професор., Универзитет у Београду – Факултет организационих наука

Контакт: boris.delibasic@fon.bg.ac.rs
Област интересовања: Системи за подршку одлучивању, Вишекритеријумско одлучивање, Пословна интелигенција, Машинско учење