

ПРИМЕНА МОДЕЛА ВРЕМЕНСКИХ СЕРИЈА ЗА ПРЕДИКЦИЈУ СРЕДЊЕГ ГЛОБАЛНОГ НИВОА МОРА**APPLICATION OF TIME SERIES MODEL FOR GLOBAL MEAN SEA LEVEL PREDICTION**Стефан Арађанин, Јелена Сливка, *Факултет техничких наука, Нови Сад***Област – РАЧУНАРСТВО И АУТОМАТИКА**

Кратак садржај – Глобално загревање се односи на стално и дугорочно повећање просечне површинске температуре Земље. То је првенствено узроковано људским активностима као што су сагоревање фосилних горива, крчење шума и испуштање штетних гасова у атмосферу, попут угљен-диоксида и метана. Глобално загревање има дубоке ефекте на океане и мора широм света. Пораст нивоа мора резултира разним последицама на Земљи и захтева хитне мере за ублажавања. Овај рад истражује како различити фактори утичу на месечни пораст нивоа мора и његове временске флукуације. Разматрани су фактори попут температуре, стопе топљења глечера, густине мора, салинитета и нивоа угљен-диоксида. Први део рада фокусира се на прикупљању, *data wrangling*-у и истраживачкој анализи података (енг. *Explorative Data Analysis, EDA*) којима је заједнички атрибут временска одредница. Други део рада фокусира се на примену модела временских серија (енг. *Time Series*) коришћењем *XGB Regressor*-а, у циљу предвиђања тачне промене средњег глобалног нивоа мора. Користећи сређене и анализиране податке, модел може да укључи низ различитих фактора који утичу на ниво мора и да предвиди њихову промену на основу историјских образаца, са веома ниском стопом грешке.

Кључне речи: глобално загревање, средњи глобални ниво мора, машинско учење, временске серије

Abstract – *Global warming refers to the ongoing and long-term increase in the Earth's average surface temperature. It is primarily caused by human activities such as burning fossil fuels, deforestation, and releasing harmful gases into the atmosphere, like carbon dioxide and methane. Global warming has profound effects on oceans and seas worldwide. The rising sea levels result in various consequences on Earth and require urgent mitigation measures. This paper investigates how various factors influence monthly sea level rise and its temporal fluctuations. These factors include temperature, glacier melting rates, sea density, salinity, and carbon dioxide levels. The first part of the paper focuses on data collection, data wrangling, and exploratory data analysis*

НАПОМЕНА:

Овај рад проистекао је из мастер рада чији ментор је била др Јелена Сливка, ванр. проф.

(*EDA*), all centered around the common element - time. The second part of the paper concentrates on applying the Time Series model, specifically using the XGB Regressor, to predict precise global mean sea level changes. Using cleaned and analyzed data, the model can incorporate various factors affecting sea levels and forecast its changes based on historical patterns with a very low error rate.

Keywords: *global warming, global mean sea level, machine learning, time series*

1. УВОД

Глобално загревање представља трајни и непрекидни пораст површинске температуре Земље. Његово основно порекло лежи у људским активностима, укључујући сагоревање фосилних горива, крчење шума и испуштање значајних количина штетних гасова у атмосферу, као што су угљен-диоксид и метан. Повећање глобалних температура већ је довело до пораста нивоа мора и чешћих и озбиљнијих природних катастрофа. Штавише, глобално загревање оставља дубок траг на светске океане и мора. Како температуре расту, океанске струје се померају, а морска вода се загрева и закисељује због повећаног нивоа угљен-диоксида. Ове промене доводе до значајних поремећаја у морским екосистемима, узрокујући катастрофалне последице за бројне водене врсте.

Овај рад истражује утицај различитих фактора на месечне флукуације нивоа мора и њихове временске обрасце. Скуп података обухвата низ међусобно повезаних фактора који су од суштинске важности за глобално загревање, посебно за пораст нивоа мора. Ови чиниоци обухватају температуру, стопе топљења глечера, густину океана, салинитет, нивое угљен-диоксида и друге. Сви они су повезани заједничком временском ознаком, омогућавајући свеобухватно истраживање њихових међусобних односа и њиховог заједничког утицаја на глобално загревање.

Да би се постигао циљ прецизног предвиђања пораста нивоа мора, било је неопходно користити различите аналитичке и предиктивне методе користећи тренутно доступне податке. Сходно томе, извршено је прикупљање ограничених, али релевантних података из различитих извора који се тичу нивоа мора и фактора који су му допринели. Затим су примењене технике *data wrangling*-а за сређивање и организацију података. Детаљна анализа, користећи статистичке

методе, служи да би се идентификовали обрасци и трендови у подацима. На крају, спроведена је истраживачка анализа како би се истражиле потенцијално интересантне варијабле које доприносе променама нивоа мора.

Скуп података искоришћен је за решавање проблема временских серија који се односи на тачно превиђање средњег глобалног нивоа мора (енг. *Global Mean Sea Level, GMSL*). Упркос малом обиму података, примена анализе временских серија се показала као веома ефикасан алат за прецизно предвиђање. Користећи добро структурисане и анализирани податке, модел укључује различите факторе који утичу на пораст нивоа мора и предвиђа његове будуће промене на основу историјских образаца, са изузетно ниском маргином грешке. На пример, модел показује изузетну тачност у предвиђању пада нивоа мора, постижући апсолутну грешку од 0.03 и грешку предвиђања од 0.5 за *GMSL* између августа и септембра 2013. Вреди напоменути да модел лакше предвиђа опадање у односу на пораст нивоа мора.

У наредном поглављу ће бити представљена већ постојећа, слична решења овом раду. У поглављу 3, описана је имплементација система, док је у поглављу 4 приказана дискусија о резултатима. На крају, у последњем поглављу, представљен је закључак рада.

2. ПРЕТХОДНА РЕШЕЊА

У раду [1], вршена је реконструкција глобалног нивоа мора користећи по податке добијених мерењем плимe и осеке у периоду од 1870. до 2004. године. Резултати до којих су научници дошли указује на пораст нивоа мора и то од 195 мм и стопу пораста у 20. веку од $1,7 \pm 0,3$ мм годишње. Главни недостаци овог рада су да се рад ослања искључиво на податке о плими и осеки који лако могу бити подложни променама, као и то да нису покушали да открију допринос глечера и ледених површина на регионални образац пораста нивоа мора, што признају и сами аутори. Такође, иако је познато убрзање *GMSL*-а, оно се тешко може открити због малог броја података.

У раду [2], анализиран је раст глобалне температуре на Земљи користећи неколико главних алгоритама машинског учења над прикупљеним подацима. Модели су креирани анализом односа између температуре и фактора као што су концентрације угљен-диоксида, азот-оксида и метана. Искоришћена је линеарна интерполација како би се попуниле недостајуће вредности. Коришћена је осмострука унакрсна валидација за тражење одговарајућих хипер-параметара модела и упоређене су перформансе четири различита алгорита машинског учења: *Linear Regression, Lasso, SVR* и *Random Forest*. Као најбољи алгоритам се показао *Random Forest* и издвојио угљен-диоксид, као главни фактор који доприноси повећању температуре, а потом га прате метан и азот-оксид.

У раду [3], извршена је анализа термичке експанзије у регионима отвореног океана која доводи до промена у приобалним нивоима мора, а како би се одредила тенденција повећања/смањења нивоа мора. Коришћен је модел регресије Гаусовог процеса и рекурентна

неуронска мрежа. Скуп података датира између 1993. и 2018. године. Утврђено је да промене температуре низ водени стуб до 700 м дубине утичу на промене варијабилности регионалног обалног нивоа мора. У раду је показано да велике температурне флукуације утичу на промену нивоа мора за већину региона, ипак представљене су и мане, а то је да би се додатно могла укључити анализа кретања копна, промена опсега плимe и промена леда.

У раду [4], приказана је улога модела временске серије у процесу предвиђања уз анализу података временских серија користећи записе о месечном средњем нивоу мора од јануара 1978. до октобра 2020. године на Великом Острву у Луизијани. Коришћен је *SARIMA* модел за предвиђање будућих тачака у временској серији. Модел је показао линеарни узлазни тренд повећања нивоа мора уз константну стопу раста у краткорочном периоду. Такође, у раду је коришћена и вишеслојна перцептрон неуронска мрежа. Резултати рада указују да коришћење модела временских серија могу пружити одличан приступ за предикцију средњег глобалног нивоа мора.

3. МЕТОД

У наредним поглављима су описани коришћени скупови података, примењене технике њихове обраде ради креирања коначног скупа, као и начин креирања предиктивног модела временске серије.

3.1. Прикупљање и обрада података

Прикупљање података подразумевало је прегледање јавно доступних скупова података, прикупљених из различитих извора, а који садрже употребљива обележја и податке о факторима који утичу на пораст нивоа мора. Издвојено је пет скупова сирових података, који су потом засебно процесирани у даљим корацима.

У скупу података који садржи информације о укупној површини морског леда [6], прикупљени су подаци у периоду од 26.10.1978. до 31.5.2019. године. Иницијалном анализом података, установљено је да су мерења конзистентна само на месечном нивоу и да скуп података садржи 26 354 редова. Обележја која чине овај скуп су: датум мерења, укупна површина морског леда, недостајуће вредности мерења, извор мерења и северна и јужна Земљина сфера. Укупна површина морског леда и датум мерења су обележја од значаја.

Наредни обрађивани скуп података садржи информације о узорцима воде сакупљане из океана између 1959. и 2020. године [5]. Иницијални скуп података садржи 337 792 реда и доста дуплицираних вредности. Емпиријским путем је закључено да би обележја, чија је стандардна девијација мања од 3, требало уклонити из скупа. Неки од атрибута скупа су: датум прикупљања, дубина океана на ком је извршено мерење, концентрације кисеоника, нитрата, нитрита и силиката. Током анализе, утврђено је да велике дубине немају значај када је у питању регионални пораст нивоа мора и, по угледу на овај закључак, су узети у обзир само подаци прикупљени на дубинама до 200 метара.

У трећем скупу података [7], забележене су температуре од чак 1750. до 2015. Године. Приликом прегледања скупа, закључено је да су вредности датума мерења, просечних температура копна, као и температуре копна и океана једине од интереса у овом раду. Очишћени скуп података садржи 3 192 реда.

У четвртом раду [8], поред датума у ком је извршено мерење, налази и податак о просечној молској фракцији угљен-диоксида одређеног из дневног просека. Подаци су прикупљани од 1958. до 2018. године.

Последњи, најбитнији, скуп података [9] садржи промену нивоа мора мерену у периоду између 1880. до 2013. године и оно представља циљно обележје.

3.2. Data wrangling

Овај процес укључује технике манипулације над прикупљеним подацима, како би скуп података био што чистији и сачињен од међусобно усклађених података спремних за моделовање временских серија.

Анализом свих временских обележја, емпиријским путем је одабрано да ће се прикупљени скупови података агрегирати на основу месеца у ком је забележено мерење и да ће највећа искоришћеност прикупљених података бити уколико се одабере временска одредница од 1969. до 2013. године.

У скупу података [6] извршено је груписање података путем израчунавања средње вредности како би подаци били употребљени на основу месеца мерења. Обзиром да скуп података садржи податке од 1978. године, емпиријским путем, испробавањем неколико техника додавања недостајућих вредности, одлучено је да оне буду попуњене коришћењем *Curve Fitting* математичке функције.

У скупу података са узорцима воде [5], установљено је да постоји 3080 различитих вредности за датум, од укупно 236 431 редова. Извршено је груписање података коришћењем *median* методе и тиме су дупликати уклоњени, а подаци о температури воде, концентрацији кисеоника, силиката и нитрата су задржани за одређени дан. Због великог броја недостајућих вредности, извршена је интерполација. Емпиријским путем и испробавањем неколико типова интерполације, дошло се до закључка да полиноминална функција даје најбоље резултате.

Остали скупови података, осим уклањања датума који су ван одабраног опсега, нису захтевали детаљније трансформације.

3.3. Истраживачка анализа података

Истраживачка анализа је искоришћена за откривање образаца понашања података у сваком од испитаних скупова, уз визуелни приказ корисних обележја и њихових вредности.

Употребом неколико библиотека *python* програмског језика, вршена је провера одступајућих вредности у скупу података о површини морског леда [6]. С обзиром на то да график није приказао одступнике, долази се до закључка да скуп података садржи конзистентну дистрибуцију, без присуства екстремних вредности. Анализом скупа података о водама [5], идентификоване су тачке које значајно одступају од

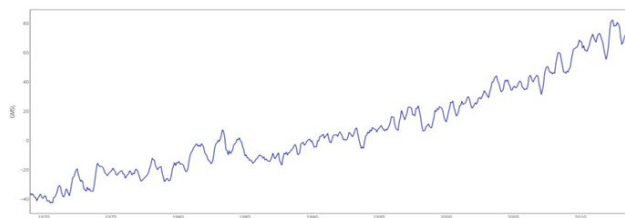
очекиване дистрибуције и ове вредности су изостављене из скупа да би се смањило лошији утицај на тачност предиктивног модела временских серија.

У свим осталим скуповима података, поновљени су исти кораци за детекцију аномалија и одступајућих вредности међу подацима и установљено је да нема екстремних вредности.

Након извршених свих корака претпроцесирања, скупови података су спремни за заједничку агрегацију на основу датума мерења. Резултујући скуп података садржи следеће атрибуте:

- *Date* – датум мерења изражен у формату година - месец,
- *Extent* - укупна површина морског леда изражена у јединици 10^6 km^2 ,
- *WaterTemp* – температура воде изражена у целзијусима,
- *O2ml* – засићеност воде кисеоником,
- *SiO3* – концентрација силиката,
- *NO3* – концентрација нитрата,
- *LandAverageTemperature* – глобална просечна температура копна изражена у целзијусима,
- *LandAndOceanAverageTemperature* - глобална просечна температура копна и океана изражена у целзијусима,
- *CO2* - просечна месечна молска фракција CO_2 и
- *GMSL* – средњи глобални ниво мора.

На слици 1, дат је графички приказ временске серије који приказује еволуцију средњег глобалног нивоа мора током времена. Уочен је тренд инкременталног пораста нивоа мора током посматраног периода који указује кретање навише, са благим сезонским осцилацијама.



Слика 1. Графички приказ временских серија *GMSL-a*

3.4. Архитектура решења

За предикцију средњег глобалног нивоа мора, коришћен је модел временских серија. *GMSL* означава циљно обележје коју модел настоји да предвиди, док се сва остала обележја користе као предиктивне карактеристике за помоћ при обучавању.

Улаз у модел додатно чине хронолошке тачке посматрања. Модел је у стању да предвиђа будуће вредности на основу претходних зависних корака. У овом раду, на основу *K* корака, предвиђа се (*K*+1)-ви корак.

ForecasterAutoreg је употребљен за предикцију модела, при чему је за регресор одабран *XGB Regressor* алгоритам. Приликом обучавања, узете вредности за *XGB Regressor* су биле [50, 100, 500] за *n_estimators*, [3, 5, 6, 10] за *max_depth* и [0.01, 0.1] за *learning_rate*. За временске кораке кашњења, модел временске серије је приликом обучавања издвојио да

низ корака [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] даје најбоље резултате.

4. РЕЗУЛТАТИ И ДИСКУСИЈА

Резултујући претпроцесирани скуп података подељен је на три подскупа: обучавајући, валидациони и тест скуп. За метрику процене учинка предвиђања средњег глобалног нивоа мора коришћена је средња апсолутна грешка (енг. *Mean Square Error*) података. Поред тога, коришћено је и тестирање уназад. У склопу тестирања уназад, урађена симулација предвиђања се одвија за сваки претходни месец са кораком један, уз процену предиктивних способности модела укључујући и тест скуп података.

Најбоље перформансе модела су показале да је вредност метрике средње апсолутне вредности грешке 70.8. Метрика тачности је приликом тестирања уназад имала вредност 56.3. Ова вредност представља средњу квадратну грешку предвиђања модела.

Анализирањем добијених резултата може се закључити да модел греши при предвиђању већих растућих вредности средњег глобалног нивоа мора. При мањим и опадајућим вредностима *GMSL*-а, модел има изузетно добре перформансе. У табели 1, приказана су три најбоља предвиђана месеца.

Датум	<i>GMSL</i>	Предвиђени <i>GMSL</i>	Апсолутна грешка
2013-08	71.6	71.57	0.03
2013-06	71.5	72	0.5
2013-07	71.2	72.4	1.2

Табела 1. Вредности најбољих перформанси модела

Модел временских серија је као факторе који највише утичу на предиктивне способности издвојио: корак временског кашњења од 1 до 10, дебљину леда, температуру воде, ниво кисеоника, ниво силиката, ниво нитрата, температуру копна и копна и океана и на крају количину угљен-диоксида. Извођењем закључка из свих сличних радова и ове студије, може се доћи до претпоставке да је таквом временом пораст температуре имао утицај на дебљину леда, а који је потом имао утицај на повећање средњег глобалног нивоа мора. Управо су време, дебљина леда и температура воде, фактори који највише доприносе крајњим резултатима излаза из модела.

5. ЗАКЉУЧАК

У овом раду представљен је систем за предвиђање будућег пораста средњег глобалног нивоа мора, уз испитивање различитих фактора који на њега утичу. Мотивација за овај рад произилази из дубоког утицаја глобалног загревања на океане и мора широм света.

На почетку је циљ био прикупити што већи број скупова података из различитих извора. Потом је извршена обрада података с обзиром на то да је постојало доста неупотребљивих вредности и мерења која не би могла да се уклопе са осталим подацима из преосталих скупова података.

Наредни корак претпроцесирања података био је *data wrangling* током ког су подаци трансформисани на начин на који би се могли агрегирати по месецу мерења и додатно су уклоњени шумови и попуњене недостајуће вредности. Истраживачка анализа података помогла је да се отклоне одступници. За предвиђање будућих промена *GMSL*-а, коришћен је модел временске серије, при чему је за регресор одабран *XGB Regressor* алгоритам.

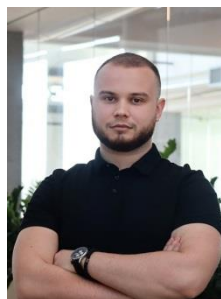
Средња апсолутна грешка приказује да модел греши са вредношћу 56.3. Најбоље предвиђани датум мерења јесте август 2013. године и за њега модел греши са вредношћу од само 0.03.

За даља унапређења, могу се истражити перформансе модела обученог над засебним скуповима података. Могуће је и поново обучити модел само са атрибутима за које се показало да највише утичу на његове предиктивне способности. Осим коришћења *XGB* регресора, требало би испитати и перформансе других алгоритама.

6. ЛИТЕРАТУРА

- [1] John A. Church, Neil J. White, 2006: A 20th century acceleration in global sea-level rise.
- [2] Harvey Zheng, vol. 7, no. 3 2018: Analysis of Global Warming Using Machine Learning.
- [3] Veronica Nieves, Christina Radin, Gustau Camps-Valls, 2021: Predicting regional coastal sea level changes with machine learning
- [4] Yeong Nain Chi, 2022: Time Series Modeling and Forecasting of Monthly Mean Sea Level (1978 – 2020): SARIMA and Multilayer Perceptron Neural Network
- [5] <https://www.kaggle.com/datasets/mathsian/water-temperature>
- [6] <https://www.kaggle.com/datasets/nsidcorg/daily-sea-ice-extent-data>
- [7] <https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data?select=GlobalTemperatures.csv>
- [8] <https://datahub.io/core/co2-ppm>
- [9] <https://datahub.io/core/sea-level-rise>

Кратка биографија:



Стефан Арађанин рођен је 1998. године у Кикинди. Основне академске студије завршио је 2021. године на Факултету техничких наука у Новом Саду, на ком брани и мастер рад 2023. године из области Електротехнике и рачунарства.
контакт: aradjani.stefan07@gmail.com