

Софтверско решење за обраду, предикцију и визуализацију података о земљотресима

A Software Solution for the Processing, Prediction and Visualization of the Earthquake Data

Нина Кузминац, Владимир Димитриески, Факултет техничких наука, Нови Сад

Студијски програм – РАЧУНАРСТВО И АУТОМАТИКА

Кратак садржај – Данас, услед све веће количине података о сеизмичким активностима, јавља се потреба за системима који ће омогућити благовремено, поуздано и разумљиво праћење и анализу земљотреса. Циљ овог рада је омогућавање анализе сеизмичких података, прогнозе магнитуде и праћења земљотреса у реалном времену уз помоћ развијеног софтверског решења. Стога, имплементирано је софтверско решење које обједињује пакетну и обраду токова података, примену машинског учења и интерактивну визуализацију ради анализе историјских података, праћења актуелне сеизмичке активности, предикције магнитуде и обавештавања о потенцијално опасним догађајима.

Кључне речи: пакетна обрада података, обрада токова података, визуализација података, предикција магнитуде земљотреса

Abstract – Nowadays, due to the increasing amount of data on seismic activities, there is a growing need for systems that will enable timely, reliable, and comprehensible monitoring and analysis of earthquakes. The aim of this work is to enable seismic data analysis, magnitude prediction and earthquake monitoring in real time with the help of a developed software solution. Therefore, a software solution was implemented that combines batch and stream processing, application of machine learning and interactive visualization for the purpose of analyzing historical data, monitoring current seismic activity, magnitude prediction and alerting about potentially dangerous events.

Keywords: batch processing, real-time processing, data visualization, earthquake magnitude prediction

НАПОМЕНА: Овај рад проистекао је из мастер рада чији ментор је био др Владимир Димитриески, ван. проф.

1. УВОД

Земљотреси представљају један од најзначајнијих природних феномена који директно угрожавају људске

животе, инфраструктуру и економску стабилност читавих региона.

Развојем глобалних мрежа сеизмолошких станица и савремених технологија за акумулацију информација, количина података о сеизмичкој активности расте из године у годину. Истовремено, повећава се и потреба за њиховим ефикасним складиштењем, обрадом и анализом.

Последњих година, све већи значај добијају технике машинског учења. Примењене на историјске сеизмичке податке, оне омогућавају изградњу модела који могу да предвиде одређене карактеристике новооткривених земљотреса или препознају обрасце који нису уочљиви класичним статистичким приступима.

Важан аспект представља и визуализација сеизмичких података, како историјских, тако и оних који пристижу у реалном времену. Разумевање механизма настанка земљотреса, као и идентификација образаца у подацима, од пресудног је значаја за унапређивање система раног упозоравања и планирање инфраструктуре отпорне на потресе.

Данас, све више људи има потребу да разуме и прати сеизмичке активности, због њиховог потенцијално опасног утицаја, што захтева брз и поуздан увид у податке, без оптерећења детаљима начина њихове обраде. Због тога постоји потреба за решењем које омогућава анализу историјских података, праћење актуелних догађаја, процену потенцијалне магнитуде у реалном времену и правовремено упозоравање у случају повећаног сеизмичког ризика, кроз интуитиван и разумљив приказ података.

Циљ овог рада је омогућавање анализе сеизмичких података, прогнозу магнитуде и праћење земљотреса у реалном времену уз помоћ развијеног софтверског решења. Комбинацијом анализе историјских података и обраде података који пристижу у реалном времену, софтверско решење представљено у овом раду настоји да обезбеди дубљи увид у обрасце појављивања земљотреса и омогући бржу процену њихове потенцијалне снаге.

2. АНАЛИЗА ПОСТОЈЕЋИХ СИСТЕМА

Анализа сеизмичких података представља кључну компоненту у праћењу сеизмичке активности, процени

ризика и унапређењу процедура заштите. У наставку поглавља представљени су неки од најзначајнијих софтверских решења који се користе у пракси за обраду података о земљотресима, анализу таласних облика и управљање мрежама сеизмолошких станица. *SeisComP* је један од најраспрострањенијих софтверских система за аквизицију, процесирање и дистрибуцију сеизмичких података у реалном времену [1]. Првенствено се користи у професионалним сеизмолошким центрима. Подржава аутоматску детекцију и локацију земљотреса, рад са више станица, транспорт података у реалном времену. Такође, омогућава стабилан рад националних и регионалних служби, што га чини стандардом у сеизмолошким институцијама.

ObsPy је *Python* радни оквир отвореног кода намењен читању, обради и анализи сеизмичких података [2]. Подржава све кључне формате сеизмичких података и интеграцију са глобалним сервисима попут *IRIS*-а, *GFZ*-а или *ORFEUS*-а. Пружа развојне алате за филтрирање сигнала, корекције инструмента, *STA/LTA* детекцију и визуализацију. *ObsPy* је кључан алат у истраживању и изради научних и оперативних апликација.

SEISAN, систем за сеизмичку анализу, је скуп програма и једноставна база података за анализу земљотреса из аналогних и дигиталних података [3]. Садржи *GUI* и служи за одређивање сеизмичког момента, локација, анализу спектра итд. Намењен је пост-сеизмичкој анализи, а често се користи у научним институцијама. Корисницима омогућава уређивање догађаја и статистичке извештаје.

USGS је глобално најпознатија платформа за праћење земљотреса, пружајући *ComCat* каталог, *API*-је, интерактивне мапе земљотреса и *ShakeMap* карте померања тла [4]. Поред тога, *USGS* развија *PAGER* систем за процене фаталних случајева и економских губитака након потреса. Платформа је референтни извор за глобална обавештења и податке у реалном времену.

EMSC је међународна организација која обједињује податке из бројних сеизмолошких мрежа и прикупљање извора путем заједнице [5]. Њихова апликација *LastQuake* користи аутоматско препознавање повећања активности корисника (посете сајту, мобилна апликација, друштвене мреже) како би детектовала потресе у реалном времену [6]. Пружа брза обавештења, атрактивне мапе, листе догађаја и пријаве примећених земљотреса од стране становништва.

GEOFON (GFZ Potsdam) корисницима пружа приступ опсежној међународној мрежи трајних сеизмолошких станица и богатој архиви записа о сеизмолошким таласима, укључујући податке у реалном времену и архивске записе, као и алате и веб-сервисе за преузимање и анализу тих података. Такође, омогућава аутоматско одређивање основних параметара земљотреса и дистрибуцију параметара догађаја научној заједници, центрима за рано упозоравање и другим релевантним институцијама.

Наведена софтверска решења и интернет платформе представљају основу савремене сеизмолошке праксе, али су углавном усмерене на прикупљање, основну

анализу и приказ података. Развој сопственог решења произилази из потребе за дубљом аналитиком, интеграцијом великих историјских скупова података и обрадом догађаја у реалном времену, као и применом техника машинског учења. Предложено софтверско решење, описано у овом раду, обједињује обраду историјских и података у реалном времену, кориснички прилагођене упите, предикцију магнитуде и интерактивну визуализацију. На тај начин не надовезује се само на постојеће изворе података, већ омогућава напредну анализу, аутоматизовано доношење закључака и благовремено информисање корисника, што га издваја као аналитички и предиктивни алат у односу на постојећа решења.

3. АРХИТЕКТУРА СОФТВЕРСКОГ РЕШЕЊА

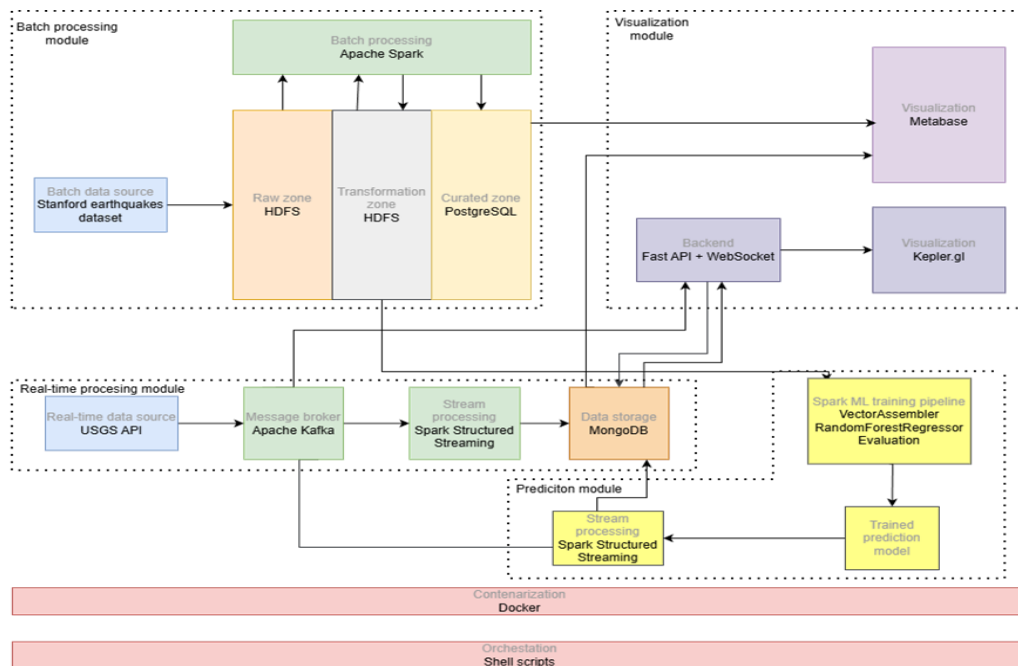
Софтверско решење за обраду, предикцију и визуализацију сеизмичких података заснива се на скупу међусобно интегрисаних компоненти које заједно формирају јединствено софтверско решење за обраду података у пакетном режиму и у режиму обраде у реалном времену, предикцију магнитуде и визуализацију резултата. Архитектура софтверског решења приказана је на слици 1 и организована је у четири функционалне целине: пакетну обраду података, обраду токова података, модул за предикцију магнитуде и визуализациони модул.

Ове целине функционишу као засебне, али повезане компоненте, које размењују податке путем заједничких складишта и комуникационих механизма. На тај начин софтверско решење омогућава да се историјски подаци, текући догађаји и предиктивни резултати интегришу у кохерентан процес континуиране анализе.

3.1. Пакетна обрада података

Прва компонента платформе представља модул за пакетну обраду података (енгл. *Batch processing module*) и представљена је у горњем левом углу на слици 1. Пакетна обрада обухвата систем за управљање великим историјским скуповима података о сеизмичкој активности. У овом делу софтверског решења врши се централизовано прикупљање, складиштење и трансформација података ради формирања стабилне аналитичке основе.

Историјски подаци преузимају се из спољних извора и смештају у *HDFS*, у оквиру сирове зоне која служи као трајни репозиторијум неизмењених оригиналних датотека. Ова зона обезбеђује накнадну реконструкцију било ког корака обраде. Над подацима се затим извршава обрада у оквиру окружења *Apache Spark*. *Spark* је задужен за скалабилну трансформацију, укључујући структурирање података, припрему временских и геолокационих атрибута и формирање аналитичких скупова. Резултати се складиште у зони трансформација *HDFS*-а, где добијају уједначен формат погодан за аналитичке упите и даљу употребу у модулу машинског учења. Коначни обрађени подаци уписују се у базу података *PostgreSQL* која представља уређену зону, који у архитектури обавља улогу централизоване аналитичке базе података.



Слика 1. Архитектура софтверског решења

3.2. Обрада токова података

Другу функционалну целину чини модул за обраду података у реалном времену (енгл. *Real-time processing module*) и представљен је испод модула за пакетну обраду података на слици 1. У оквиру овог модула, подаци се преузимају у реалном времену из *USGS API*-ја и представљају текуће сеизмичке догађаје који се глобално ажурирају у кратким интервалима. Пристижући догађаји прво се прослеђују преко платформе *Kafka*. *Kafka* омогућава хоризонталну скалабилност, високу пропусност и гарантовану испоруку порука, што софтверско решење чини отпорним на прекиде и нестабилности у спољним сервисима.

Spark Structured Streaming је интегрисан као конзумент токова *Kafka*. *Spark Structured Streaming* у архитектури обавља улогу механизма за континуирано структурирање података и њихово прилагођавање за даље коришћење у складишту и модулима предикције. Резултати добијени након извршавања упита над трансформисаним подацима се уписују у базу података *MongoDB*, која служи као складиште полуструктурираних података у реалном времену. Ова компонента архитектуре омогућава да софтверско решење реагује на нове догађаје у року од неколико секунди, што је неопходно за апликације чија је сврха праћење сеизмичке активности у реалном времену.

3.3. Визуализација података

Визуализациони модул (енгл. *Visualization module*), приказан у горњем десном углу на слици 1, представља кориснички оријентисан део архитектуре и интегрише резултате пакетне обраде и обраде токова података у јединствен интерфејс. Овај модул обухвата два комплементарна подсистема: аналитичку визуализацију историјских података и динамичку геопросторну визуализацију догађаја у реалном

времену. Аналитички део је заснован на платформи *Metabase*, која се ослања на *PostgreSQL* као извор података. У овој компоненти користе се агрегирани и историјски подаци настали током пакетне обраде, уз могућност формирања контролних табли, статистичких прегледа и временских анализа. Други подсистем реализује визуализацију у реалном времену путем алата *Kepler.gl*, односно праћење актуелне сеизмичке активности путем интерактивне мапе. Визуализација се заснива на подацима о земљотресима који тренутно пристижу у реалном времену, као и на подацима који су раније примљени у реалном времену и били сачувани у бази података *MongoDB*. Подаци које серверски слој прими у реалном времену чувају се у бази података како би обезбедили историјски преглед догађаја насталих у реалном времену, што омогућава да корисник приликом отварања *Kepler.gl* мапе има комплетан увид у раније и актуелне земљотресе, без приказа празне мапе.

Комбинацијом ова два визуализациона механизма софтверско решење пружа истовремено увид у дугорочне трендове и тренутна дешавања, чинећи визуализацију кључним делом целокупне архитектуре.

3.4. Модул за предикцију магнитуде

Модул за предикцију магнитуде (енгл. *Prediction module*), приказан у доњем десном углу на слици 1, представља интелигентни део архитектуре који се ослања на машинско учење ради предикције магнитуде земљотреса. Он је структуриран као комбинација процеса обучавања и процеса извођења предикције. Обучавање се изводи периодично, где библиотека *Spark ML* користи припремљене историјске податке за изградњу модела. Одабране колоне се комбиновањем у вектор припремају за улаз у модел користећи *VectorAssembler*, чиме се испуњавају формални услови библиотеке *Spark ML*. На тај начин се свака

појединачна евиденција своди на структуру погодну за обучавање и тестирање. За предикцију је изабран модел *RandomForestRegressor* јер омогућава учење нелинеарних односа између атрибута и добро подноси шум и нестандартну расподелу података типичну за сеизмичке догађаје.

Након завршетка обучавања, модел се процењује помоћу метрика *RMSE*, *MAE* и R^2 . Ове мере омогућавају да се оцени колико добро модел реконструише стварну магнитуду и у којој мери објашњава варијансу присутну у подацима. Ниске вредности *RMSE* и *MAE*, добијене приликом евалуације модела, указују на добру тачност и стабилност предикција магнитуде земљотреса, док вредност коефицијента детерминације R^2 показује да модел објашњава већину варијабилности у подацима. Обучени модел се затим чува у централизованом складишту модела које је доступно осталим компонентама софтверског решења.

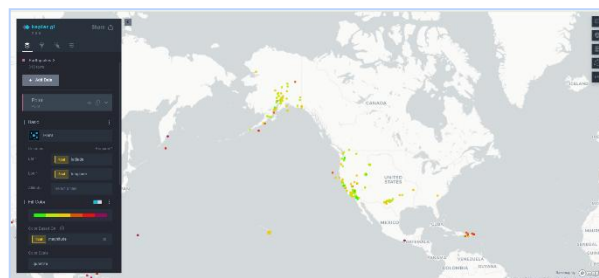
У реалном времену, модул за предикцију магнитуде учитава већ обучени модел и примењује га на нове догађаје који пристижу кроз платформу *Kafka*. Предикције се затим уписују у базу података *MongoDB* и постају доступне визуализационим и обавештавајућим подсистемима. Овај модул такође обухвата логику за генерисање упозоравајућих сигнала када предикција пређе унапред дефинисан праг.

Оваква организација омогућава да софтверско решење комбинује историјске податке, догађаје у реалном времену и машинско учење у јединствен, аутоматизован и скалабилан предиктивни механизам.

4. ЗАКЉУЧАК

Примарни циљ развијеног софтверског решења је омогућавање анализе сеизмичких података, прогнозе магнитуде и праћења земљотреса у реалном времену. Софтверско решење обезбеђује интегрисану обраду историјских и текућих података, као и упозоравање у случају потенцијално опасних сеизмичких активности. Коришћењем платформи *Spark* и *Kafka* и компоненте *Spark Structured Streaming* омогућено је континуирано прикупљање, обрада и трансформација података, уз скалабилну и флексибилну архитектуру. Модел машинског учења, обучен над историјским подацима, примењује се над новим догађајима у реалном времену, уз сталну процену тачности предикције.

У случају прекорачења дефинисаног прага прогнозиране магнитуде, систем шаље обавештење путем електронске поште. Резултати пакетне обраде приказују се кроз *Metabase*, док се подаци који пристижу у реалном времену визуализују на мапи у алату *Kepler.gl*, као што је приказано на слици 2. Сваки земљотрес је приказан као тачка на мапи, где различите боје указују на разлике у магнитуди и највећа концентрација земљотреса јавља се дуж тектонских плоча. Резултати постигнути током развоја софтверског решења показали су да је могуће интегрисати пакетну обраду и обраду токова података са моделом машинског учења и обезбедити платформу која у реалном времену обрађује земљотресе, предвиђа њихову магнитуду и пружа корисне и правовремене информације кроз визуализационе алате.



Слика 2. Визуализација земљотреса који се догађају у реалном времену уз помоћ *Kepler.gl*

Иако софтверско решење показује висок ниво функционалности и стабилности, постоје ограничења која треба имати у виду. Прецизност модела машинског учења зависи од квалитета и обима историјских података, док кашњења у долазним догађајима могу утицати на правовременост предвиђања. Такође, инфраструктурни захтеви окружења *Spark* могу бити значајни у условима велике фреквенције података.

Потенцијална проширења софтверског решења укључују увођење напреднијих модела машинског учења, побољшани *feature engineering* и интеграцију података из више сеизмичких извора. Такође, софтверско решење се може проширити механизмима за геопросторну анализу, предвиђањем накнадних потреса и интеграцијом са постојећим платформама за цивилну заштиту, чиме би се унапредио квалитет раног упозоравања.

5. ЛИТЕРАТУРА

- [1] "SeisComP seismological software." [Online]. Available: <https://www.seiscomp.de/>. [Accessed: 01-Dec-2025].
- [2] "ObsPy Documentation (1.4.2)." [Online]. Available: <https://docs.obspy.org/>. [Accessed: 01-Dec-2025].
- [3] "SEISAN." [Online]. Available: <https://seisan.info/>. [Accessed: 01-Dec-2025].
- [4] "Earthquake Hazards Program | U.S. Geological Survey." [Online]. Available: <https://www.usgs.gov/programs/earthquake-hazards/>. [Accessed: 01-Dec-2025].
- [5] "EMSC – European-Mediterranean Seismological Centre." [Online]. Available: <https://www.emsc-csem.org/>. [Accessed: 01-Dec-2025].
- [6] "LastQuake." [Online]. Available: <https://m.emsc.eu/>. [Accessed: 01-Dec-2025].

Кратка биографија:



Нина Кузминац, рођена је 21. јануара 2002. године. Након завршених основних академских студија, уписала се на мастер академске студије на Факултету техничких наука, студијски програм Рачунарство и аутоматика.

Контакт: kuzminacn@gmail.com