

Примена предиктивне аналитике у анализи саобраћајних незгода у Београду путем Power BI алата

Predictive Analytics Approach to Traffic Accident Analysis in Belgrade Using Power BI

Сандра Лазаревић, Факултет техничких наука, Нови Сад

Студијски програм – ИНЖЕЊЕРСТВО ИНФОРМАЦИОНИХ СИСТЕМА

Кратак садржај – Рад приказује примену предиктивне аналитике у анализи саобраћајних незгода у Београду за период 2020–2025. Полазни подаци, преузети са платформе отворених података Републике Србије, обрађени су у Python-у и Power BI-ју. Применом техника чишћења, моделовања и визуализације развијени су регресиони, класификациони и кластер модели ради предвиђања ризичних периода и локација. Резултати су приказани кроз интерактивни Power BI dashboard са временским трендовима, географским расподелама и прогнозама броја незгода. Анализа указује на сезонске и просторне обрасце који могу допринети унапређењу безбедности саобраћаја.

Кључне речи: предиктивна аналитика, отворени подаци, саобраћајне незгоде, Power BI, машинско учење

Abstract – The paper presents the application of predictive analytics in analyzing traffic accidents in Belgrade for the period 2020–2025. The dataset, obtained from the official Open Data Portal of the Republic of Serbia, was processed using Python and Power BI. Through data cleaning, modeling, and visualization, regression, classification, and clustering models were developed to predict high-risk periods and locations. Results are displayed in an interactive Power BI dashboard showing temporal trends, geographic distributions, and accident forecasts. The analysis reveals seasonal and spatial patterns that can support improvements in urban traffic safety.

Keywords: predictive analytics, open data, traffic accidents, Power BI, machine learning

НАПОМЕНА: Овај рад проистекао је из мастер рада чији ментор је доц. др Теодора Вучковић

1. УВОД

Безбедност саобраћаја у Београду представља значајан изазов због велике густине саобраћаја и броја возила. Упркос унапређењима, број незгода остаје висок, што указује на потребу за савременијим методама анализе. Рад представља интегрисани систем за анализу и предикцију саобраћајних незгода заснован на Python-

у, техникама машинског учења и Power BI-ју. Користећи податке са портала data.gov.rs [1], примењени су регресиони, класификациони и кластер модели ради предвиђања броја незгода и идентификације ризичних локација. Резултати су визуализовани кроз интерактивне извештаје који омогућавају ефикасније управљање безбедношћу саобраћаја.

2. МЕТОДОЛОГИЈА ИСТРАЖИВАЊА

Истраживање је спроведено кроз више међусобно повезаних фаза које обухватају прикупљање, припрему и обраду отворених података, примену метода машинског учења и визуелну интерпретацију резултата у Power BI окружењу. Посебан акценат стављен је на интеграцију аналитичких и визуелних компоненти, чиме је омогућено обједињено сагледавање временских, просторних и узрочних аспеката саобраћајних незгода. Комбинацијом програмског језика Python и алата Power BI, развијен је истраживачки оквир који повезује статистичко моделовање, предиктивно учење и визуелну аналитику.

Такав приступ омогућава да се комплексни подаци о незгодама преточе у разумљиве, интерактивне и практично применљиве информације које могу допринети унапређењу система безбедности саобраћаја.

2.1. Извори података

Подаци коришћени у истраживању преузети су са званичног портала data.gov.rs [1], из скупа „Подаци о саобраћајним незгодама по полицијским управама и општинама“ [2]. Сет података обухвата период 2020–2025. и више од 80.000 записа који садрже: датум, време, локацију (управа, општина), тип и исход незгоде (материјална штета, повређени, погинули). Фокус истраживања је на територији града Београда, а подаци су јавно доступни, анонимизовани и редовно ажурирани, што омогућава њихову поуздану примену у истраживачке сврхе.

2.2. Припрема и обрада података

У првој фази извршено је чишћење и трансформација података у Python окружењу, уз следеће кораке:

- уклањање дупликата и неконзистентних вредности,
- конверзија временских поља у стандардизовани формат,
- кодирање категоријских променљивих нормализација нумеричких вредности.

Након тога, подаци су припремљени за моделовање и подељени на тренинг и тест скупе (однос 80:20), чиме је омогућена објективна процена перформанси модела.

2.3. Примена модела машинског учења

За предиктивну анализу примењени су следећи алгоритми:

- Linear Regression – за предвиђање броја незгода по месецима и општинама,
- Decision Tree Classifier – за класификацију тежине незгода (материјална штета, повређени, погинули),
- K-Means алгоритам – за идентификацију „црних тачака“, односно локација са високом концентрацијом незгода.

Перформансе модела процењиване су помоћу MAE, RMSE, R^2 и F1-мере, а затим су резултати експортирани у Power BI за визуелну анализу и интерпретацију.

2.4. Интеграција и визуализација

Добијени резултати повезани су у Power BI кроз релационе моделе (димензије: датум, општина, исход). Визуализације обухватају:

- линијски дијаграм за праћење сезонских трендова,
- просторну мапу за приказ “црних тачака”,
- класификационе приказе по исходу незгода и
- упоредни приказ предвиђених и стварних вредности.

Овакв систем омогућава интерактивну анализу и динамично сагледавање безбедности саобраћаја у Београду, уз могућност проширења модела новим подацима и додатним метрикама.

3. ИМПЛЕМЕНТАЦИЈА И РЕЗУЛТАТИ АНАЛИЗЕ

Процес имплементације обухватио је примену модела машинског учења у Python окружењу и њихову интеграцију у Power BI, ради визуелне интерпретације и анализе резултата. На тај начин обезбеђен је потпуни циклус обраде — од прикупљања и моделовања података, до интерактивног приказа кључних налаза.

3.1. Припрема података и окружење

Анализа је спроведена у окружењу Google Colab, коришћењем библиотека pandas, numpy, scikit-learn и matplotlib. Након филтрирања података на полицијску управу Београд, извршено је чишћење и стандардизација dataset-a, као и издвајање релевантних атрибута — „општина“, „датум_незгоде“, „исход_незгоде“, „тип_незгоде“, „лонгитуда“, „латитуда“. На основу колоне датума креиране су нове променљиве („година“, „месец“,

„дан“, „дан_у_недељи“), чиме је омогућена временска сегментација и анализа сезонских образаца. Основна анализа показала је да се највећи број незгода дешава у периоду од маја до октобра, најчешће петком, а просторно су најзаступљеније општине Нови Београд, Вождовац и Палилула.

3.2. Резултати модела машинског учења

За предвиђање месечног броја незгода коришћен је Linear Regression алгоритам.

Модел је показао просечну грешку (MAE) од око 110 незгода месечно, што представља одступање од 7–10% у односу на стварне податке. Коефицијент детерминације ($R^2 = -0.13$) указује на слабу линеарну везу између времена и броја незгода, што потврђује да је модел користан за илустрацију тренда, али не и за прецизне пројекције.

На основу графика 1, може се приметити да модел уопштено препознаје тренд раста и пада броја незгода, али уз уочљива одступања у појединим периодима. Ово потврђује да је линеарна регресија адекватна за илустрацију основних трендова, али не и за прецизно предвиђање.

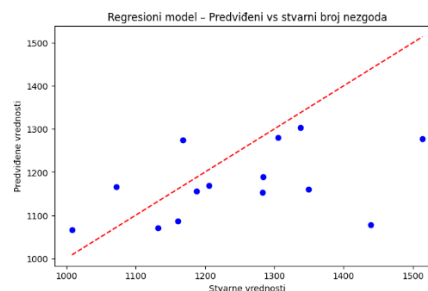


График 1. Регресиони модел - Поређење стварних и предвиђених вредности броја незгода по месецима

За предвиђање типа саобраћајне незгоде примењен је Decision Tree Classifier, који омогућава интерпретабилан приказ одлука.

Модел је трениран на атрибутима „општина“, „месец“ и „дан“, а циљна променљива је „исход_незгоде“ (материјална штета, повређени, погинули).

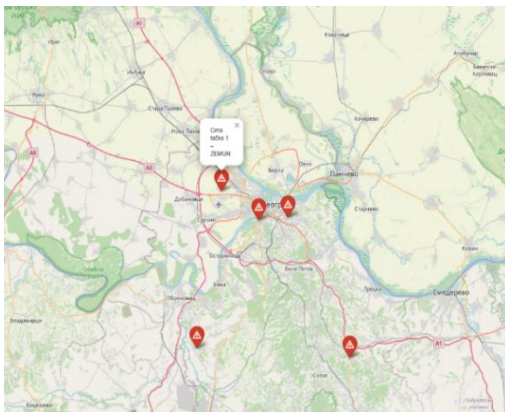
Резултати евалуације показују да модел најтачније препознаје незгоде са материјалном штетом (прецизност=0.74, одзив=1.00), док су остале класе мање заступљене и самим тим теже за предвиђање.

Анализа стабла одлучивања показала је да сезонски фактори имају највећи утицај на тежину последица, а одређене општине се издвајају као подручја са већим уделом тежих незгода.

За просторну анализу коришћен је K-Means алгоритам, који је омогућио груписање незгода према географским координатама.

Формирано је пет кластера, чије локације могу да се виде на слици 1, који означавају подручја са највећом учесталашћу саобраћајних незгода. Највећа густина забележена је у централним и јужним деловима града, посебно на територији Вождовца, Чукарице и Звездаре, док су мањи кластери примећени на периферним саобраћајницама.

Овакав приступ омогућава идентификацију ризичних зона које могу бити приоритет у планирању безбедносних мера и унапређењу инфраструктуре.



Слика 1. Мапа Београда са приказом црних тачака саобраћајних незгода

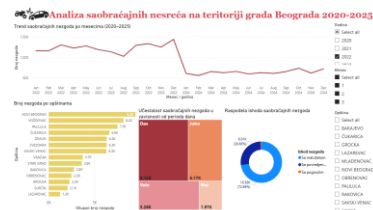
3.3. Интеграција резултата у Power BI

Power BI је изабран као алат за завршну фазу истраживања јер омогућава повезивање, обраду и визуализацију података у интерактивном окружењу. Након што су у Python-у спроведене фазе чишћења, трансформације и моделовања података, резултати су извезени у CSV формату и интегрисани у Power BI ради визуелне анализе и интерпретације.

У Power BI окружењу израђен је интерактивни *dashboard* који обједињено приказује резултате дескриптивне, предиктивне и просторне анализе. Кључне компоненте укључују:

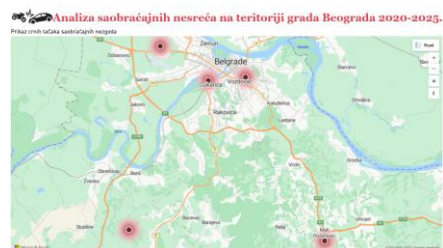
- Линијски дијаграм – приказује сезонске трендове броја незгода по месецима и годинама, што омогућава идентификацију периода са повећаним ризиком.
- Мапа црних тачака (енгл. *Heat Map*) – визуализује просторну дистрибуцију незгода, при чему су најинтензивније зоне уочене у општинама Вождовац, Чукарица и Звездара.
- Стубичасти дијаграм по општинама – омогућава поређење укупног броја незгода на територији Београда.
- Кружни дијаграм (енгл. *Pie Chart*) – приказује структуру незгода према исходу (материјална штета, повређени, погинули).
- Предиктивни приказ – пореди стварне и предвиђене вредности броја незгода, чиме се визуелно потврђује ефикасност регресионог модела и уочава очекивани пад броја инцидената у наредним годинама.

Почетна страница извештаја која се види на слици 2, приказује основне статистичке показатеље и временске трендове броја саобраћајних незгода. Овај део *dashboard*-а пружа кориснику брзи преглед укупног стања и динамичких промена кроз време.



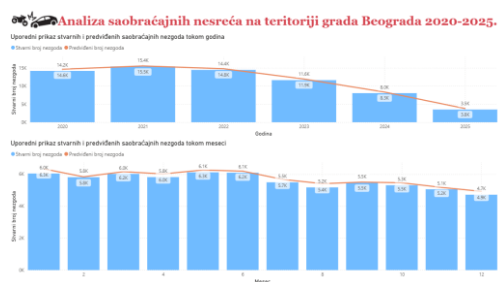
Слика 2. Почетна страница извештаја

Друга страница извештаја која је на слици 3 фокусира се на просторну димензију анализе – мапа црних тачака визуализује подручја са највећом учесталошћу незгода што омогућава идентификацију „критичних зона“.



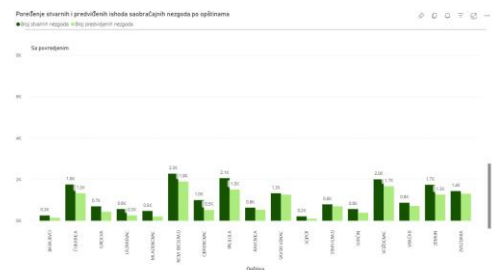
Слика 3. Друга страница извештаја

Трећа страница извештаја на слици 4 садржи стубичасти дијаграм који приказује укупни број незгода по општинама на територији Београда.



Слика 4. Трећа страница извештаја

Четврта страница извештаја на слици 5 приказује резултате предиктивне анализе. Посебно су истакнуте општине и периоди са најизраженијим разликама између стварних и очекиваних вредности, што доприноси бољем разумевању фактора ризика и трендова безбедности.



Слика 5. Четврта страница извештаја

3.4. Дискусија и правци будућих истраживања

Спроведена анализа показала је да примена предиктивне аналитике омогућава ефикасно препознавање просторних и временских образаца саобраћајних незгода у Београду. Комбинацијом Python моделовања и Power BI визуализације развијен

је интегрисани систем који омогућава увид у сезоналност, концентрацију и очекиване трендове незгода.

Регресиони модел делимично објашњава сезонске промене, али није довољан за дугорочна предвиђања. Класификациони модел постигао је тачност од 0.74 за незгоде са материјалном штетом, док су тежи исходи теже предвидиви због неуравнотежености података. Кластер анализа је идентификовала „црне тачке“ — зоне највећег ризика, углавном дуж главних раскрсница и магистралних праваца. Главна ограничења односе се на непотпун скуп података (недостају подаци о времену, стању пута и возача) и немогућност Power BI-а да динамички ажурира моделе у реалном времену. Будућа истраживања требало би да укључе напредније алгоритме (нпр. Random Forest, Gradient Boosting, неуронске мреже), технике балансирања класа (SMOTE) и додатне изворе података, као и развој интерактивних *dashboard*-а у реалном времену. Такав приступ представља основу за развој паметних система безбедности у оквиру концепта *Smart City*.

4. ЗАКЉУЧАК

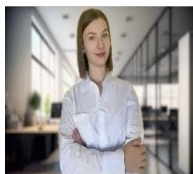
Истраживање је обухватило развој предиктивног модела за анализу саобраћајних незгода у Београду кроз интеграцију Python-а и Power BI-а. Овим приступом обезбеђен је увид у временске и просторне обрасце ризика, као и могућност предвиђања трендова. Регресиони модел успешно је уочио сезоналне варијације, класификациони је омогућио препознавање типова незгода, а кластер анализа идентификовала критичне зоне. Визуелизација у Power BI-ју допринела је бољем разумевању налаза и њиховој практичној примени.

Рад потврђује да комбинација предиктивне аналитике и визуелних алата може подржати планирање превентивних мера и унапредити систем безбедности саобраћаја у урбаним срединама.

5. ЛИТЕРАТУРА

- [1] Република Србија, *Портал отворених података* – *Data.gov.rs*, Доступно на: <https://data.gov.rs/sr/datasets>
- [2] Република Србија, Министарство унутрашњих послова, *Подаци о саобраћајним незгодама по полицијским управама и општинама*, Доступно на: <https://data.gov.rs/sr/datasets/podatsi-o-saobratshajnim-nezgodama-po-politsijskim-upravama-i-opshtinama/>

Кратка биографија:



Сандра Лазаревић, рођена је 13. фебруара 2001. у Пријепољу. 2019. године уписала је Факултет техничких наука у Новом Саду, студијски програм Инжењерство информационих система. Звање дипломираног инжењера стекла је 2023. године, а мастер студије

наставља у истом усмерењу, уз даљи професионални развој у области пројектовања и развоја база података.

Контакт: sandra.lazarevic123@gmail.com