



Sveučilište u Zagrebu

Fakultet organizacije i informatike

Saša Mitrović

**Metodološki okvir za izradu analitičkog
informacijskog sustava temeljenog na
metodama strojnog učenja u
produksijskom okružju**

DOKTORSKI RAD

Varaždin, 2025.



Sveučilište u Zagrebu

Fakultet organizacije i informatike

SAŠA MITROVIĆ

**Metodološki okvir za izradu analitičkog
informacijskog sustava temeljenog na
metodama strojnog učenja u
produksijskom okružju**

DOKTORSKI RAD

Mentori:

prof. dr. sc. Neven Vrčec

dr. sc. Valentina Janev

Varaždin, 2025.



University of Zagreb

Faculty of Organization and Informatics

Saša Mitrović

**Methodological framework for developing
analytical information system based on
machine learning methods in a production
environment**

DOCTORAL THESIS

Supervisors:

Neven Vrčec, PhD, full professor

Valentina Janev, PhD

Varaždin, 2025.

SAŽETAK

Rad predstavlja razvijeni automatizirani algoritam latencije temeljen na selekciji atributa skupa podataka koji skraćuje vrijeme predviđanja kod korištenja klasifikacijskih ili regresijskih metoda strojnog učenja u produkcijskom okružju. Glavni čimbenici razvijenog algoritma smanjenja latencije predviđanja su broj atributa u skupovima podataka, vrste atributa u skupovima podataka, postavke klasifikacijskih ili regresijskih metoda strojnog učenja, izvlačenje najznačajnijih atributa iz skupova podataka te broj predviđanja koji je moguće ostvariti u određenom vremenskom razdoblju – što je važna mjera u produkcijskom okružju. U radu, osim automatiziranog algoritma latencije, predstavljena je prilagođena verzija CRISP-DM metodologije za analitički informacijski sustav, odnosno za pretprocesiranje i obradu podataka s algoritmom latencije u analitičkom informacijskom sustavu. Rad odstupa od tradicionalnog vrednovanja modela strojnog učenja isključivo s mjerom postotka greške ili točnosti jer u velikom broju stvarnih situacija u produkcijskom okružju jedan posto razlike u točnosti predviđanja između različitih metoda strojnog učenja nema preveliki utjecaj ako se predviđanje može ostvariti u značajnije kraćem vremenu.

Ključne riječi: algoritam, automatizirano strojno učenje, automatizirani cjevovod, redukcija dimenzionalnosti, metodološki okvir

ABSTRACT

The paper will propose the development of an automated latency algorithm based on the selection of data set attributes that will shorten the prediction time when using classification/regression machine learning methods in a production environment. The main factors of the proposed prediction latency reduction algorithm are number of attributes in datasets, attribute types in datasets, settings of classification/regression machine learning methods, extraction of the most important attributes from datasets and number of predictions that can be achieved in a certain timespan - an important measure in the production environment. In addition to the automated latency algorithm, the paper will propose the development of a customized version of the CRISP-DM methodology for analytical information system, namely for pre-processing and data processing with the prediction latency algorithm in analytical information system. The paper seeks to move away from evaluating machine learning models solely with a measure of error or accuracy because in many real-world situations the one percent difference in prediction accuracy between different machine learning methods does not have much effect if the prediction can be achieved in a significantly shorter time.

Keywords: Algorithm, Automated machine learning, Automated pipeline, Dimensionality reduction, Methodological framework