UNIVERZA NA PRIMORSKEM

PEDAGOŠKA FAKULTETA KOPER

Edukacijske vede, 3. stopnja

ALGORITMI STROJNEGA UČENJA ZA IZBIRO UČNE METODE TANDEMSKEGA UČENJA NA PODROČJU DIDAKTIKE MATEMATIKE

Dispozicija doktorske disertacije

Mentor: Kandidat:

prof. dr. Darjo Felda Bor Bregant

Somentor

dr. Daniel Doz

Koper, 2025

Navedba znanstvenega področja

Doktorska disertacija v svojem jedru združuje interdisciplinarno znanje iz pedagogike, didaktike matematike ter področij rudarjenja podatkov in strojnega učenja, s ciljem izboljšanja izbire učne metode v tandemu pri poučevanju matematike. Z uporabo metod umetne inteligence, zlasti klasifikacije in izbire ključnih značilk, raziskava preučuje kompleksne vzorce med interakcijo učiteljev in učencev ter njihov vpliv na učne rezultate, ki diktirajo, ali je posamezniku obravnavana učna metoda ustrezna ali ne. Študija ponuja dragocen vpogled v prilagodljive pristope k personalizirani pedagogiki ter uporabo tehnologije za izboljšanje učnega procesa v matematičnem izobraževanju.

ARRS klasifikacija: 5.01.01, 2.07.08

MSC 2020 klasifikacija: 97D40, 97D60, 62P99

Izvleček

200-250 besed

**Ključne besede:**

Vzgoja in izobraževanje, tandemsko učenje, strojno učenje, podatkovno rudarjenje.

Teoretična izhodišča raziskave

Opredelitev problema, ciljev in hipotez disertacije oziroma raziskovalnih vprašanj ter pričakovanih rezultatov

Prispevek k znanosti

This PhD dissertation focuses on the integration of machine learning techniques to assess student responses within the context of tandem learning in high school mathematics. By leveraging a diverse array of variables, the study aims to ascertain their significance in gauging student engagement and progress. Employing methods such as Mutual Information (MI) and Recursive Feature Elimination (RFE), the research identifies the pivotal factors influencing student outcomes. The core methodology involves a three-state classification model, utilizing Likert scale responses, and employing a suite of supervised machine learning models for predictive analysis. The dissertation's impact lies in its multifaceted approach, offering insights into the nuanced interplay between learning methodologies, student variables, and predictive modeling within the educational landscape. Its findings hold the potential to inform educators and policymakers about the effectiveness of tandem learning in mathematics education and pave the way for tailored, data-driven approaches to enhance student learning experiences.

Predvidene metode raziskovanja

Struktura oziroma kazalo predvidene vsebine disertacije

Navedba osnovne literature in virov

# Uvod

300.000 do 600.000 znakov s presledki

# Teoretični del

## Pouk in učne oblike

### Kratko o pouku

…

### Neposredna in posredna učna oblika

Blazic, Kramar, Kubale, Tomic, Puklek, Marentic Pozarnik, Peklaj

Blazic, Kramar, Tubale, Tomic, Puklek, Marentic Pozarnik, Peklaj - literatura o tandemskem (skupinskem) delu. Uporabi se Slavin in roschelle (dobr\_clanek\_tandem) več!

## Delo v tandemu

### Začetki, razvoj in možnosti dela v tandemu

...

### Potek in struktura dela v tandemu

...

### Prednosti in slabosti dela v tandemu

...

## Strojno učenje in klasifikacija

### Uvod v umetno inteligenco

A diagram of a plane

Description automatically generatedUmetna inteligeca (AI) je sposobnost digitalnega računalnika ali računalniško vodenega robota, da opravlja dela pogosto asociirana z inteligentnimi bitji (Copeland, 2023). Pogosto imamo v mislih projekte izdelovanja sistemov, ki uporabljajo procese vezane na karakteristiko ljudi, kot je sposobnost razuma, iskanja smisla in učenja iz preteklih izkušenj. Ali je sistem inteligenten, lahko preverimo s Turingovim testom. V tem testu izpraševalec postavlja vprašanja, vendar ne ve, ali je na drugi strani človek ali računalnik. Računalnik opravi test, če izpraševalec po odgovorih ne more ugotoviti, ali je na drugi strani človek ali računalnik. Ena od največjih slabosti testa je, da je subjektiven in ga ni mogoče ponoviti ali matematično analizirati.

Figure 1: Turingov test: Vir slike:https://en.wikipedia.org/wiki/Turing\_test v APA stilu

### Strojno učenje

Strojno učenje je podpomenka umetne inteligence, kjer na podatkih izvajamo določene algoritme (citat iz nekje). Gre za avtomatsko detekcijo ustreznih in pomembnih vzorcev v naboru podatkov. Ker pa je v zadnjih desetletjih postala obdelava podatkov ključnega pomena za skoraj vsako nalogo, je strojno učenje postalo eden od temeljev informacijske tehnologije in s tem centralni, toda pogosto skriti del naše družbe (Vishwanathan & Smola, 2008). Strojno učenje nas obkroža na vsakem koraku: Od brskalnikov, ki nam skušajo prikazati relevantne vsebine (in s tem podtakniti personalizirane oglase), filtriranje neželene pošte in vzpostavljanja varnosti pri bančnih transakcijah. Digitalne kamere se učijo prepoznavati obraze, mikrofoni se učijo prepoznati zvočne ukaze. Avtomobili so opremljeni s tehnologijo, ki s pomočjo strojnega učenja preprečuje nesreče. V vseh teh primerih opazimo kompleksnost samih vzorcev in same situacije, kjer bi človek (programer) sam težko našel eksplicitne rešitve, kako rešiti nalogo (Shalev-Shwartz & Ben-David, 2014).

Let us begin our mathematical analysis by showing how successful learning can be achieved in a relatively simplified setting. Imagine you have just arrived in some small Pacific island. You soon find out that papayas are a significant ingredient in the local diet. However, you have never before tasted papayas. You have to learn how to predict whether a papaya you see in the market is tasty or not. First, you need to decide which features of a papaya your prediction should be based on. On the basis of your previous experience with other fruits, you decide to use two features: the papaya’s color, ranging from dark green, through orange and red to dark brown, and the papaya’s softness, ranging from rock hard to mushy. Your input for figuring out your prediction rule is a sample of papayas that you have examined for color and softness and then tasted and found out whether they were tasty or not. Let us analyze this task as a demonstration of the considerations involved in learning problems. Our first step is to describe a formal model aimed to capture such learning tasks.

Model v istem viru

(Shalev-Shwartz & Ben-David, 2014)

### Naloga klasfikacije in izbire napovednih spremenljivk

.../splošno o nalogi/

#### Modeli uporabljeni za klasifikacijo

/opis modelov in psevdoalgoritmi/

#### Izbrira napovednih spremenljivk

/kratko na sploh in o algoritmih MI in RFE/

#### Evalvacija modelov strojnega učenja

/o metrikah accuracy, precision, recall, F1/

### Naloga reduciranja dimenzije prostore

V zgoraj opisanih nalogah je pogosto smiselno reducirati dimenzijo prostora spremenljivk in ga projecirati na manjši prostor. Razlogov za to je več. Strojno učenje je pogosto računalniško zahteven proces. Kompleksnost modela hitro (pogosto eksponentno) narašča z dimenzijo prostora, kar povzroči časovno ne-ekonomičnost računalnika, ki si jo pogosto ne moremo privoščiti (na primer primer preprečevanja avtomobilske nesreče). Manjšanje dimenzije lahko tudi olajša interpretacijo podatkov, ki nam pomaga najti smiselne strukture in problem tudi vizualizirati (Shalev-Shwartz & Ben-David, 2014).

# Empirični del

## Raziskovalni problem, namen in cilji

V doktorski disertaciji smo celovito analizirali kolaborativno delo v pedagoškem okolju. Identificirali smo problem, ki izvira iz zadržkov, ki jih učitelji čutijo glede uporabe skupinskega pristopa pri pouku. Naš prispevek temelji na novem konceptualnem okvirju, ki izkorišča napredke v strojnem učenju. Naš glavni cilj je razviti model, ki bo omogočal napovedovanje učinkovitosti skupinskega dela na posamezne udeležence. To nameravamo doseči z identifikacijo ključnih spremenljivk, ki vplivajo na uspešnost skupinskega učenja.

Namen doktorske disertacije je raziskati učinkovitost algoritmov strojnega učenja pri napovedovanju odziva učenca v tandemskem učnem okolju pri matematiki v srednji šoli. Z uporabo tehnik klasifikacije in ocene pomembnosti značilk študija si prizadeva raziskati zapleten odnos med različnimi dejavniki in učnim uspehom učenca v takšnem okolju. S poglobljenim analiziranjem ta raziskava prizadeva razkrinkati, kako modeli strojnega učenja lahko natančno napovedujejo prilagodljivost in uspeh učenca v tandemskih učnih situacijah ter tako ponudijo vpogled v personalizirane izobraževalne strategije.

Cilji disertacije obsegajo večplastne naloge. Prvenstveno si prizadeva natančno oceniti najpomembnejše značilke v tandemskem učnem okolju, ki bistveno vplivajo na uspeh učenca. To vključuje celovito preučevanje različnih dejavnikov, s ciljem prepoznati njihov vpliv na učne rezultate. Poleg tega raziskava skuša identificirati najbolj učinkovite algoritme strojnega učenja za napovedno modeliranje v tem kontekstu ter oceniti njihovo natančnost, robustnost in prilagodljivost. Poleg tega je ključno tudi primerjanje učinkovitosti izbranega algoritma z alternativnimi metodami, kar omogoča celovito razumevanje njihovih prednosti in omejitev pri napovedovanju odziva učencev v tandemskih učnih okoljih matematike srednje šole. V končni fazi si ti cilji prizadevajo zagotoviti operativne uvide za izobraževalce in odločevalce, kako izboljšati učinkovitost tandemskih učnih okolij.

## Raziskovalne hipoteze

Na podlagi pregleda obstoječe literature o strojnem učenju v izobraževalnih okoljih smo oblikovali hipoteze glede na cilje raziskave.

Splošna raziskovalna hipoteza:

H: Strojno učenje omogoča celostnejši opis dijakovega odnosa do skupinskega dela z analizo večplastnih dejavnikov, kar lahko vodi v bolj natančno napovedovanje njihovega odziva v takšnih učnih okoljih.

Specifične raziskovalne hipoteze:

H1: Mnenja učencev na delo v skupini se razlikujejo glede na: ...

H2: Nekatere metode bolje napovejo dijakov odziv do skupinskega učenja kot druge

## Metodologija

Za potrjevanje raziskovalnih hipotez bomo uporabili kavzalno-(ne?)eksperimentalno metodo pedagoške raziskave. Deskriptivna metoda bo namenjena opisu vzorca, interferenčna metoda pa se bo posluževala matematičnih metod strojnega učenja.

### Vzorec

Po predpripravi podatkov, je finalna kohorta obsegala n dijakov s 54 rešenimi odgovori, ki so določali m spremenljivk. Osebe vključene v raziskavo so bili dijaki drugega in tretjega letnika neke gimnazije v Ljubljani v šolskem letu 2023/24. Vzorec je bil neslučajnostni in namenski. Deskriptivno statistiko vzorca opisujejo tabele ... in slike ...

A diagram of different feelings

Description automatically generated

A graph with a bar and a number

Description automatically generated with medium confidence

### Zbiranje podatkov

Po pridobitvi informiranih soglasij dijakov in odobritve ravnatelja šole, vključene v raziskavo smo zbrali in preučili uspeh tandemskega učenja glede na več spremenljivk. Uspeh (na splošno glede učenja in diverzifikacije pouka) smo merili v treh stanjih (dobro, nevtralno in slabo). Neodvisne spremenljivke so bile splošne narave (spol, razred, profesor in prejšnji uspeh pri matematiki), psihološke narave (MBTI spremenljivke: ekstravertnost-introvertnost, intuicija-zaznavanje, čutenje-mišljenje in presojanje-opazovanje ter druge spremenljivke: matematična tesnoba in motivacija) ter glede na timsko učenje (kakovostna interakcija, količinska interakcija in ali je študent presegel svojega partnerja). Podatki so bili anonimizirani z uporabo kodne sheme, tako da sta bila anonimnost in objektivnost zagotovljeni v vsakem koraku raziskave. Zbrane podatke je imel dostop le raziskovalec.

Podatki so bili zbrani po tem, ko so bili dijaki, vključeni v raziskavo, vključeni v okolje tandemskega učenja v obdobju približno enega tedna. Del pouka je bil namenjen običajnemu delu v učilnici, del pa delu v tandemu. Naključnost ni bila upoštevana. Dijaki so bili razporejeni v pare glede na njihovega partnerja na dvosedežni mizi.

Vsi udeleženci so sodelovali prostovoljno in niso bili finančno nagrajeni za sodelovanje v raziskavi. Raziskava je potekala v skladu z etičnimi standardi Deklaracije iz Helsinkov iz leta 1964 in evropskim zakonom o varstvu podatkov (Splošna uredba o varstvu podatkov EU–GDPR UE 2016/67).

Za osebnostne (psihološke) spremenljivke smo uporabili test MBTI, natančneje Odprte razširjene jungovske tipološke lestvice (OEJTS) kot odprtokodno alternativo. Vprašanja so bila zbrana iz (*Fastest Myers-Briggs Test*, n.d.), ki temelji na (*Myers-Briggs/Jung Test: Open Extended Jungian Type Scales*, n.d.), oba pa sta na voljo za javno uporabo pod licenco Creative Commons. Test MBTI ima tako zagovornike (Carlson, 1985; Carlyn, 1977; Randall et al., 2017) kot tudi nasprotovanja (Boyle, 1995; Coan, 1978; Druckman & Bjork, 1991). Njegova veljavnost in zanesljivost morata biti upoštevani kot previdnostni ukrep. Test za določanje motivacije je bil pridobljen iz (Sundre et al., 2012), medtem ko je bil test za matematično tesnobo (test AMAS) pridobljen iz (*PsyToolkit*, n.d.). Testi AMAS in motivacije so bili dokazano zanesljivi, veljavni in učinkoviti v izobraževalnem kontekstu (Fiorella et al., 2021; Hopko et al., 2003; Sundre et al., 2012; Yavuz et al., 2012). Vse zgoraj navedene spremenljivke so bile obravnavane kot kontinuirane (zvezne) spremenljivke, ne kot kategorične (npr. rezultat »26« za ekstraverta namesto »ekstravert«), da bi preprečili predpostavke o bipolarnosti ljudi (Ramsay et al., 2000). To lahko privede tudi do večje natančnosti modela (Carlson, 1985; Carlyn, 1977; DeVito, 1985). Anketa je uporabljala uveljavljene elemente z manjšimi prilagoditvami, da bi se prilagodila različnim kulturnim in socialnim kontekstom, pri čemer so bili ohranjeni konstrukti instrumenta.

Celoten vprašalnik, ki je bil razdeljen s pomočjo Google forms je dostopen v Priloge.

### Obdelava podatkov

Podatki, zbrani med raziskavo, so bili analizirani z uporabo programskega jezika Python (verzija 3.11.4), predvsem s knjižnicama pandas (verzija 2.1.3) in scikit-learn (verzija 1.3.2). Nepredelani anonimizirani nabor podatkov skupaj s statistično kodo vseh analiz je dostopen na ...

Preliminarna analiza bo modificirala nabor podatkov v obliko tidydata (Wickham, 2014). Taki podatki se nanašajo na strukturirano obliko, kjer vsaka spremenljivka zaseda svoj stolpec, vsak opazovanec je v svoji vrstici, različne enote opazovanja pa so organizirane v ločene tabele. Ta organizacija poenostavi obdelavo, analizo in vizualizacijo podatkov ter olajša postopek z minimalnim naborom orodij za upravljanje raznolikih in zapletenih zbirk podatkov (Wickham, 2014). Dobljenemu naboru podatkov smo nato določili podatkovni tip (npr. razred je kategorična spremenljivka, starost pa določa celo število). Kategorične spremenljivke smo nato s pomočjo “označevalnega enkodiranja” pretvorili v številke (npr. razredu 2. c priredimo število 4), kar omogoča strojem, da bolje razume in obdeluje podatke. Vnosi, ki ne bodo popolni (npr. manjka ena vrednost) bomo iz nabora podatkov izbrisali. Podatkov ne bomo skalirali, saj želimo, da so posamezne komponente disertacije primerljive in uporabne z bodočimi raziskavami, ki bodo uporabljale enake instrumente.

Za ugotavljanje veljavnosti prve hipoteze tj. katere spremenljivke najbolj vplivajo na dijakov odziv na tandemsko delo se bomo poslužili Mutual information in Recursive feature elimination.

Drugo hipotezo tj. kateri algoritmi najbolje napovejo odziv dijaka na delo v skupini bomo preverili z več algoritmi strojnega učenja. Evalvirali jih bomo na podlagi f\_1 metrike s pomočjo križne evalvacije 5x2.

## Rezultati in interpretacija

### Test zanesljivosti

Uporabljen vprašalnik je bil sestavljen iz več že preverjenih instrumentov. Ti so pogosto sestavljeni iz več podvprašanj, ki na podlagi več odgovorov sestavijo skupaj eno spremenljivko. Ker želimo, da so vprašanja, ki merijo enako stvar med sabo skladna, smo uporabili test notranje konsistentnosti. Omenjeni testi se razlikujejo glede na tip spremenljivke, ker pa naš vprašalnik zajemajo le trije instrumenti (ki sicer določajo šest spremenljivk), ki merijo le spremenljivke na zveznem nivoju, se lahko poslužimo le enega tipa. Izbrali smo Kronbach alpha, saj ..... V tabeli ... najdemo njegove vrednosti za posamezne pod-teste in njihove 95% intervale zaupanja. Anksioznost in motivacija imata dobro notranjo konsistentnost, kot lahko sklepamo iz literature. Na drugi strani pa imamo MBTI dimenzije, katerih konsistentnost je stvar debate. Na našem vzorcu je notranja konsistentnost za introverzijo in presojanje dobra, medtem ko je dimenzija čutenja slabo konsistentna, dimenzija intuicije pa nesprejemljiva.

|  |  |  |  |
| --- | --- | --- | --- |
| Spremenljivka | Število elementov (vprašanj) | Kronbach alpha | 95% interval zaupanja |
| Matematična anksioznost | 9 | 0.77 | [0.68, 0.83] |
| Matematična motivacija | 7 | 0.87 | [0.83, 0.91] |
| Introverzija | 8 | 0.69 | [0.58, 0.78] |
| Intuicija | 8 | 0.47 | [0.28, 0.62] |
| Čutenje | 8 | 0.54 | [0.39, 0.68] |
| Presojanje | 8 | 0.71 | [0.60, 0.79 ] |

### Test normalnosti

Normalnost podatkov smo preverili s Shapiro-Wilkovim (SW) testom, ki temelji na regresiji, saj so nekatere raziskave pokazale, da ima test v povprečju večjo statistično moč od ostalih testov (Hernandez, 2021; Khatun, 2021). Ta korak niti ni bil bistven, saj naše izbrane statistike, tako glede pomembnosti značilk, kot predikcije ne zahtevajo normalizacije. Poleg tega so bile nekatere spremenljivke v našem naboru podatkov same po sebi kategorične, kar dodatno poveže ne-pomembnost tega koraka (Rado et al., 2019; Tavazzi et al., 2020). Vse testirane spremenljivke so videti normalne. Njihove priložene Gaussove krivulje, QQ-diagrame in p-vrednosti SW statistike najdemo v slikah in tabeli.

A graph and diagram of a graph

Description automatically generated with medium confidence

A graph and diagram of a graph

Description automatically generated with medium confidence

A graph and diagram of a graph

Description automatically generated with medium confidence

A graph and a diagram

Description automatically generated with medium confidence

A comparison of a graph

Description automatically generated

A graph and diagram of a graph

Description automatically generated with medium confidence

|  |  |
| --- | --- |
| Zvezna spremenljivka | p-vrednost SW testa |
| Anksioznost | 0.46 |
| Motivacija | 0.50 |
| Introverzija | 0.32 |
| Intuicija | 0.64 |
| Čutenje | 0.05 |
| Presojanje | 0.22 |

### Test korelacije značilk

Raziskovanje korelacije zagotavlja dragocene vpoglede v odnose med spremenljivkami, kar pomaga pri identifikaciji potencialnih povezav in odvisnosti, ki so ključne za razumevanje medsebojnega delovanja in morebitnega vpliva med različnimi dejavniki znotraj podatkovnega nabora (Patil & Franken, 2021). Poudariti je treba, da zagotavljamo le matriko korelacij, vendar pri izbiri značilk ne upoštevamo morebitnih povezav med spremenljivkami.A diagram of a person's personality

Description automatically generated with medium confidence

Figure 3: Zamenjaj za slovenska imena

Interpretacija rezultata....

Posamezni testi...

Najpogosteje uporabljeni meri korelacije sta Spearmanov in Pearsonov koeficient in oba zavzemata možne vrednosti med -1 in 1. Slednji meri linearno zvezo med dvema zveznima normalno porazdeljenima spremenljivkama, medtem ko prvi meri kakršnokoli monotono interakcijo med spremenljivkama (De Winter et al., 2016; Schober et al., 2018) in je bolj robusten za osamelce. Korelacijski koeficient k je pogosto tretiran kot zelo močen za |k|> 0.7, srednje močen za 0.5 < |k| < 0.7, znaten za 0.3 < |k| < 0.5 in šibek za |k| < 0.3 (Rovetta, 2020). Negativne vrednosti koeficianta implicirajo negativno korelacijo, medtem ko k = 0 pomeni nobene linearne zveze med spremenljivkama.

### Test pomembnosti značilk

Seznam pomembnosti z njihovimi ocenami MI in RFE so podani v Tabeli x in prikazani na Sliki x in Sliki x. Višja kot je ocena MI in nižji kot je rang RFE, večja je odvisnost med spremenljivko in izidom. Spremenljivke smo le rangirali in nismo izbrali s pomočjo statističnih testov, ali so pomembne ali ne. Kljub ne-optimalni notranji konsistentnosti in relativno nizki pomembnosti nekaterih značilk, smo vključili vse, saj bi lahko še vedno imele napovedno moč (Chen et al., 2020). Ta odločitev je bila statistično korektna, saj naš nabor podatkov ni bil majhen, zato preprileganje ni bila primarna skrb (Ying, 2019). Moramo pa imeti to v mislih, ko interpretiramo model.

A graph with different colored bars

Description automatically generated

A graph of a bar graph

Description automatically generated

|  |  |  |
| --- | --- | --- |
| Spremenljivka | MI rezultat (višja vrednost pomeni višji vpliv) | RFE rang (nižja cifra pomeni višji vpliv) |
| Outperforming partner | 0.22 | 1 |
| Razred | 0.09 | 5 |
| Kvalitativna interakcija | 0.08 | 1 |
| Profesor | 0.04 | 3 |
| Anksioznost | 0.01 | 8 |
| Spol | 0.01 | 1 |
| Ocena | 0.00 | 4 |
| Kvalitativna interakcija | 0.00 | 2 |
| Motivacija | 0.00 | 11 |
| Introverzija | 0.00 | 10 |
| Presojanje | 0.00 | 6 |
| Čutenje | 0.00 | 7 |
| Presojanje | 0.00 | 9 |

### Test predikcije

Z ozirom na oceno F1, so trije najboljši modeli v našem eksperimentu Random forest, K-Nearest Neighbors in Gradient boosting. Njihova uspešnost je srednja, ker pa je njihova natančnost večja od števila statističnih enot v največjem stratumu (49.4 %), so modeli boljši od ugibanja. Drugi modeli so bili slabše uspešni, kar implicira, da se niso dobro naučili globjih vzorcev. Celotna klasifikacija je dostopna v Tabeli x.

Table 1:Klasifikacija z vsem.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Klasifikator | Accuracy (natančnost) | Precision (preciznost) | Recall (priklic) | F1-score (ocena F1) |
| Naive Bayes | 0.338 | 0.272 | 0.377 | 0.258 |
| K-Nearest Neighbors | 0.505 | 0.344 | 0.361 | 0.343 |
| Decision Tree | 0.471 | 0.344 | 0.337 | 0.312 |
| Gaussian Mixture | 0.023 | 0.017 | 0.050 | 0.024 |
| Linear Discriminant Analysis | 0.450 | 0.328 | 0.381 | 0.311 |
| AdaBoost | 0.484 | 0.307 | 0.350 | 0.316 |
| Gradient Boosting | 0.505 | 0.367 | 0.368 | 0.341 |
| Support Vector Machine | 0.450 | 0.159 | 0.304 | 0.208 |
| Random Forest | 0.550 | 0.427 | 0.397 | 0.351 |

Za vizualizacijo uspešnosti modelov, je na Sliki x matrika zmede, ki nudi vpogled v interakcijo pravih pozitivnih, napačnih pozitivnih, pravih negativnih in napačnih negativnih ugibanj posameznega modela (Chicco et al., 2021).

Slika 2: Matrike zmede.

A group of blue squares

Description automatically generated

Da bi preprečili neuravnoteženost nabora podatkov, smo se odločili tudi za binarno klasifikacijo, tako da smo združili stratuma 0 (metoda skupnega dela mi sploh ni všeč) in 1 (do metode skupnega dela sem indiferenten) S tem smo dobili nova stratuma 0 s 45 (50.6 %) in 1 (49.4 %), kar kreira bolj ravnovesno reprezentacijo odnosa do metode skupnega dela (Harangi et al., 2020). S tem pristopom, so bili vsi algoritmi, izvzemši AdaBoost uspešni. Najbolje se je odrezal algoritem SVM z dobro natančnostjo (60.7 %) in srednjo oceno F1 (0.59).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Classifier | Accuracy | Precision | Recall | F1-score |
| Naive Bayes | 0.583 | 0.539 | 0.581 | 0.539 |
| K-Nearest Neighbors | 0.516 | 0.521 | 0.518 | 0.506 |
| Decision Tree | 0.518 | 0.511 | 0.517 | 0.504 |
| Gaussian Mixture | 0.540 | 0.581 | 0.532 | 0.447 |
| Linear Discriminant Analysis | 0.563 | 0.482 | 0.561 | 0.504 |
| AdaBoost | 0.473 | 0.421 | 0.469 | 0.420 |
| Gradient Boosting | 0.551 | 0.503 | 0.549 | 0.509 |
| Support Vector Machine | 0.607 | 0.638 | 0.607 | 0.586 |
| Random Forest | 0.561 | 0.513 | 0.560 | 0.520 |

Poslužili smo se tudi algoritmov z uporabo manj napovednih spremenljivk. Izpustili smo vseh šest spremenljivk vezanih na psihološki profil dijaka in dobili slab uspeh v vseh modelih, kar kaže, da imajo te spremenljivke vseeno nekaj napovedne moči, kot obravnavano v Poglavju x.

*Table 4: Classification results with selected features*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Classifier | Accuracy | Precision | Recall | F1-score |
| Naive Bayes | 0.339 | 0.260 | 0.376 | 0.250 |
| K-Nearest Neighbors | 0.494 | 0.343 | 0.359 | 0.340 |
| Decision Tree | 0.449 | 0.330 | 0.323 | 0.310 |
| Gaussian Mixture | 0.033 | 0.016 | 0.142 | 0.027 |
| Linear Discriminant Analysis | 0.394 | 0.239 | 0.340 | 0.251 |
| AdaBoost | 0.359 | 0.252 | 0.263 | 0.250 |
| Gradient Boosting | 0.405 | 0.255 | 0.294 | 0.265 |
| Support Vector Machine | 0.495 | 0.165 | 0.333 | 0.221 |
| Random Forest | 0.494 | 0.301 | 0.359 | 0.315 |

### Test manjšanja dimenzije prostora oziroma vizualizacija podatkov

Zaradi suboptimalne uspešnosti metod naših strojno naučenih modelov, smo se poslužili še t-SNE analize, da bi prikazali morebitne (ne)-vzorce in odnose med spremenljivkami, ki bi morda lahko vplivali na uspešnost modelov (Bibal et al., 2023). Sliki x in x kažeta, da so vzorci v naboru podatkov težki za ločitev, kot smo špekulirali.

A diagram of a graph

Description automatically generated with medium confidence

A diagram of a graph

Description automatically generated

# Sklepne ugotovitve

…

# Literatura in viri

Boyle, G. J. (1995). Myers-Briggs Type Indicator (MBTI): Some Psychometric Limitations. *Australian Psychologist*, *30*(1), 71–74. https://doi.org/10.1111/j.1742-9544.1995.tb01750.x

Carlson, J. G. (1985). Recent Assessments of the Myers-Briggs Type Indicator. *Journal of Personality Assessment*, *49*(4), 356–365. https://doi.org/10.1207/s15327752jpa4904\_3

Carlyn, M. (1977). An Assessment of the Myers-Briggs Type Indicator. *Journal of Personality Assessment*, *41*(5), 461–473. https://doi.org/10.1207/s15327752jpa4105\_2

Chen, R.-C., Dewi, C., Huang, S.-W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, *7*(1), 52. https://doi.org/10.1186/s40537-020-00327-4

Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*, *14*(1), 13. https://doi.org/10.1186/s13040-021-00244-z

Coan, R. W. (1978). *The Eighth Mental Measurements Yearbook*, *1*, 970–975.

Copeland, J. (2023). Artificial intelligence. In *Encyclopedia Britannica*. https://www.britannica.com/technology/artificial-intelligence

De Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, *21*(3), 273–290. https://doi.org/10.1037/met0000079

DeVito, A. J. (1985). *Review of the Myers-Briggs Type Indicator*. *1*, 1030–1032.

Druckman, D., & Bjork, R. A. (1991). *In the Mind’s Eye: Enhancing Human Performance* (p. 1580). National Academies Press. https://doi.org/10.17226/1580

*Fastest Myers-Briggs test*. (n.d.). Retrieved 21 October 2023, from https://dynomight.net/mbti/

Fiorella, L., Yoon, S. Y., Atit, K., Power, J. R., Panther, G., Sorby, S., Uttal, D. H., & Veurink, N. (2021). Validation of the Mathematics Motivation Questionnaire (MMQ) for secondary school students. *International Journal of STEM Education*, *8*(1), 52. https://doi.org/10.1186/s40594-021-00307-x

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, *3*(null), 1157–1182.

Hernandez, H. (2021). *Testing for Normality: What is the Best Method?* https://doi.org/10.13140/RG.2.2.13926.14406

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education. Promise and Implications for Teaching and Learning.*

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The Abbreviated Math Anxiety Scale (AMAS): Construction, Validity, and Reliability. *Assessment*, *10*(2), 178–182. https://doi.org/10.1177/1073191103010002008

Khatun, N. (2021). Applications of Normality Test in Statistical Analysis. *Open Journal of Statistics*, *11*(01), 113–122. https://doi.org/10.4236/ojs.2021.111006

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, *97*(1–2), 273–324. https://doi.org/10.1016/S0004-3702(97)00043-X

Lal, T. N., Chapelle, O., Weston, J., & Elisseeff, A. (2006). Embedded Methods. In I. Guyon, M. Nikravesh, S. Gunn, & L. A. Zadeh (Eds.), *Feature Extraction: Foundations and Applications* (pp. 137–165). Springer. https://doi.org/10.1007/978-3-540-35488-8\_6

*Myers-Briggs/Jung Test: Open Extended Jungian Type Scales*. (n.d.). Retrieved 21 October 2023, from https://openpsychometrics.org/tests/OEJTS/

Patil, V. H., & Franken, F. H. (2021). Visualization of statistically significant correlation coefficients from a correlation matrix: A call for a change in practice. *Journal of Marketing Analytics*, *9*(4), 286–297. https://doi.org/10.1057/s41270-021-00120-z

*PsyToolkit*. (n.d.). Retrieved 4 November 2023, from https://www.psytoolkit.org/index.html

Rado, O., Ali, N., Sani, H. M., Idris, A., & Neagu, D. (2019). Performance Analysis of Feature Selection Methods for Classification of Healthcare Datasets. In K. Arai, R. Bhatia, & S. Kapoor (Eds.), *Intelligent Computing* (Vol. 997, pp. 929–938). Springer International Publishing. https://doi.org/10.1007/978-3-030-22871-2\_66

Ramsay, A., Hanlon, D., & Smith, D. (2000). The association between cognitive style and accounting students’ preference for cooperative learning: An empirical investigation. *Journal of Accounting Education*, *18*(3), 215–228. https://doi.org/10.1016/S0748-5751(00)00018-X

Randall, K., Isaacson, M., & Ciro, C. (2017). Validity and Reliability of the Myers-Briggs Personality Type Indicator: A Systematic Review and Meta-analysis. *Journal of Best Practices in Health Professions Diversity*, *10*(1), 1–27.

Rovetta, A. (2020). Raiders of the Lost Correlation: A Guide on Using Pearson and Spearman Coefficients to Detect Hidden Correlations in Medical Sciences. *Cureus*, *12*(12). https://doi.org/10.7759/cureus.11794

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.

Sundre, D., Barry, C., Gynnild, V., & Tangen Ostgard, E. (2012). Motivation for Achievement and Attitudes toward Mathematics Instruction in a Required Calculus Course at the Norwegian University of Science and Technology. *Numeracy*, *5*(1). https://doi.org/10.5038/1936-4660.5.1.4

Tavazzi, E., Daberdaku, S., Vasta, R., Calvo, A., Chiò, A., & Di Camillo, B. (2020). Exploiting mutual information for the imputation of static and dynamic mixed-type clinical data with an adaptive k-nearest neighbours approach. *BMC Medical Informatics and Decision Making*, *20*(S5), 174. https://doi.org/10.1186/s12911-020-01166-2

Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing and Applications*, *24*(1), 175–186. https://doi.org/10.1007/s00521-013-1368-0

Vishwanathan, S. V. N., & Smola, A. (2008). *Introduction to Machine Learning*. Cambridge University Press. https://alex.smola.org/drafts/thebook.pdf

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*, 1–23. https://doi.org/10.18637/jss.v059.i10

Wlodzislaw, D., Winiarski, T., Biesiada, J., & Kachel, A. (2003). *Feature selection and ranking filters*.

Yavuz, G., Ozyildirim, F., & Dogan, N. (2012). Mathematics Motivation Scale: A Validity and Reliability. *Procedia - Social and Behavioral Sciences*, *46*, 1633–1638. https://doi.org/10.1016/j.sbspro.2012.05.352

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*, 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

# Priloga A: Vprašalnik

|  |  |  |  |
| --- | --- | --- | --- |
| Ciljna spremenljivka:  Označi, kje na linearni skali od 1 (Ne – zdi se mi izguba časa) do 3 (Metoda je bila zabavna in koristna) | 1. | Se ti je metoda (delo v skupini) zdela na sploh uspešna? Upoštevaj tako vidik napredovanja pri matematiki, kot popestritve pouka. |  |
| Kviz osebnosti (motivacija): Označi, koliko od 1 do 5 se strinjaš s trditvijo | 2. | V moji izobraževalni poti želim imeti čim več matematike | 2-3+4+5+6+7-8 |
| 3. | Na fakulteti bi se rad izognil matematki |
| 4. | Težavnost matematike me privlači |
| 5. | Učenje (napredne) matematike smatram za uporabno |
| 6. | Deljenje idej za reševanje matematičnega problema mi je v coni udobja |
| 7 | Rad imam matematiko |
| 8 | Matematika je dolgočasna |
| Kviz osebnosti (matematična anksioznost): Od 1 (skoraj nič anksioznosti) do 5 (velika anksioznost) označi, koliko ti sledeča stvar povzroča anksioznosti (nelagodja, tesnobe) | 9 | Uporaba in iskanje formul ter tabel na zadnji strani poglavja v učbeniku | 9+10+11+12+13+14+15+16+17 |
| 10 | Razmišljanje o testu matematike dan prej |
| 11 | Gledanje profesorja, ki na tablo rešuje enačbo |
| 12 | Pisanje testa matematike |
| 13 | Prejemanje domače naloge pri matematiki |
| 14 | Poslušanje ure matematike v razredu |
| 15 | Poslušanje sošolca, ki razlaga snov pri matematiki |
| 16 | Pisanje kratkega nenapovedanega preverjanja pri matematiki |
| 17 | Začetek nove snovi pri pouku matematike |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Kviz osebnosti (MBTI): Za vsak par označi, kje na lestvici (od 1 do 5) med elementoma, meniš, da si | 18 | Naredim seznam | Zanašam se na spomin | Introvertiranost / ekstrovertiranost:  Sensing / intuicija  Meja za vse: 24 |
| 19 | Skeptičen | Želim verjeti |
| 20 | Dolgčas mi je biti sam | Potrebujem čas zase |
| 21 | Sprejmem stvari take kot so | Nisem zadovoljen s trenutnim stanjem |
| 22 | Skrbim za čistočo sobe | Odlagam stvari kjerkoli |
| 23 | Oznaka človeka kot “robota” je žaljiva | Stremim k “mehaničnemu” razmišljanju |
| 24 | Poln energije | Relaksiran |
| 25 | Raje imam vprašanja zaprtega tipa | Raje imam esejska vprašanja |
| 26 | Sem kaotičen | Sem organiziran |
| 27 | Enostavno me je čustveno raniti | Imam trdo kožo |
| 28 | Najbolje delam v skupinah | Najbolje delam sam |
| 29 | Osredotočen sem na sedanjost | Osredotočen sem na prihodnost |
| 30 | Načrte delam vnaprej | Narte delam zadnji hip |
| 31 | Želim spoštovanje drugih | Želim ljubezen drugih |
| 32 | Zabave me utrujajo | Na zabavah zaživim |
| 33 | Fits in | Izstopam |
| 34 | Imam rezervne načrte | Dejanju sem predan |
| 35 | Želim biti dober v popravljanju stvari | Želim biti dober v “popravljanju” oseb |
| 36 | Več govorim | Več poslušam |
| 37 | Ko razalagm dogodek, povem, kaj se je zgodilo | Ko razalagm dogodek se osredotočim na njegov pomen |
| 38 | Dela se hitro lotim | Z delom odlašam (prokrastiniram) |
| 39 | Sledim srcu | Sledim glavi |
| 40 | Ostanem doma | Grem ven |
| 41 | Želim širšo sliko zgodbe | Želim podrobnosti zgodbe |
| 42 |  |  |
| 42 | Improviziram | Se pripravim |
| 43 | Etiko temeljim na pravici | Etiko temeljim na sočustvovanju |
| 44 | Ne maram kričati | kričanje, ko so poslušalci daleč mi je naravno |
| 45 | Raje imam teorijo | Raje imam prakso (empirijo) |
|  | 46 | Trdo delam | Trdo se zabavam |  |
|  | 47 | Čustva me spravijo iz cone udobja | Čustva upoštevam in cenim |  |
|  | 48 | Rad nastopam pred ljudmi | Javnemu govoru se izogibam |  |
|  | 49 | Želim vedeti kdo, kaj, kdaj. | Želim vedeti zakaj |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Splošna vprašanja | 50 | 48 | Lanska zaključna ocena pri matematiki | Možne vrednosti 1 – 5 |
| 51 | 49 | Razred | n možnih izbir |
| 52 | 50 | Spol | 2 možni izbiri |
| 53 | 51 | Učeči profesor | n možnih izbir |
| Vprašanja, ki se navezujejo na tandemsko delo | 54 | 52 | Oceni, koliko je bilo v povprečju v tandemu interakcije | 3 možne izbire:  0 (skoraj nič) – 2 (veliko) |
| 55 | 53 | Oceni, kako produktivna je bila interakcija | 3 možne izbire:  0 (več klepetanja kot produktivnega dela) – 2... |
| 56 | 54 | Si pripomogel več kot tvoj partner v tandemu | 3 možne izbire:  0 (pripomogel sem manj) – 2 (pripomogel sem več) |

# Feature selection

Feature selection methods are typically categorized into three primary groups: wrappers, embedded methods, and filter methods (Guyon & Elisseeff, 2003).

Wrappers involve the incorporation of an inductive learning algorithm as part of the process of evaluating different feature subsets (Kohavi & John, 1997). These methods commonly assess performance based on the classification rate achieved on a testing set. While wrappers may indeed yield strong generalization results, they come with the notable drawback of substantial computational demands, especially when applied to high-dimensional datasets. Furthermore, they are susceptible to issues such as overlearning and sensitivity to initialization, which can limit their practicality.

Embedded methods take a different approach by integrating knowledge about the specific structure of the class of functions employed by a given learning machine (Lal et al., 2006). Embedded methods, in comparison to wrappers, tend to be less computationally intensive. However, they remain considerably slower than filter methods and are often intertwined with the characteristics of the learning machine, thus making the selected features contingent on the specific algorithm employed.

Filter methods operate on the premise of total independence between the learning machine and the data, utilizing a metric that is agnostic to the induction learning algorithm for the assessment of feature subsets (Wlodzislaw et al., 2003). Filter methods, unlike wrappers, exhibit a degree of robustness against overfitting. However, they may not consistently identify the most optimal feature subset for classification or regression tasks, potentially leading to suboptimal results.

Each of these three feature selection methods comes with its own set of advantages and limitations, and the choice of which to employ in a given context should be guided by the specific requirements and constraints of the problem at hand. (Guyon & Elisseeff, 2003; Vergara & Estévez, 2014).