UNIVERZA NA PRIMORSKEM

PEDAGOŠKA FAKULTETA

DOKTORSKA DISERTACIJA

BOR BREGANT

KOPER 202X

UNIVERZA NA PRIMORSKEM

PEDAGOŠKA FAKULTETA

Doktorski študijski program tretje stopnje

Edukacijske vede

Doktorska disertacija

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

Bor Bregant

Koper 202x

Mentor:

prof. dr. Darjo Felda

Somentor:

dr. Daniel Doz

ZAHVALA

Želim izraziti iskreno zahvalo vsem, ki so neposredno ali posredno prispevali k uspešnemu zaključku te doktorske disertacije. Najprej bi se rad zahvalil mentorju dr. Darju Feldi za brezmejno podporo, strokovno usmerjanje in nesebično mentorstvo. Vaša modrost in predanost sta bila ključnega pomena pri mojem akademskem razvoju.

Prav tako bi se rad zahvalil somentorju dr. Danielu Dozu za njegovo stalno vodstvo in mentorstvo od samega začetka mojega študija. Vaše usmerjanje in podpora sta močno prispevala k mojemu razumevanju raziskovalnega procesa in vključitvi v raziskovalno sfero.

Posebna zahvala gre tudi dijakom in profesorjem, ki so sodelovali pri raziskavi in omogočili zbiranje podatkov ter analizo. Brez vaše dragocene udeležbe in prispevkov ne bi bilo mogoče doseči končnih rezultatov. Vaša vnema in zavzetost sta bila ključnega pomena pri razširitvi obzorij te študije.

**IZJAVA O AVTORSTVU**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Podpisani/a | Bor Bregant | | vpisna številka | | 98233001 | |
| izjavljam, da je doktorska disertacija z naslovom | | | Algoritmi strojnega učenja za izbiro | | | |
| učne metode tandemskega učenja na področju didaktike matematike, | | | | | | |
| pod mentorstvom | | prof. dr. Dario Felda | | in somentorstvom | | dr. Daniel Doz |

* rezultat lastnega raziskovalnega dela,
* da so rezultati korektno navedeni,
* da nisem kršil/a avtorskih pravic in intelektualne lastnine drugih in
* da je elektronska različica, ki sem jo oddal/a, istovetna tiskani različici.

Izjavljam, da za potrebe arhiviranja dovoljujem / ne dovoljujem (ustrezno obkrožite) objavo elektronske različice v repozitoriju Dissertations and Thesis (Proquest) in dLib.si (NUK). V skladu s 1. odstavkom 21. člena Zakona o avtorski in sorodnih pravicah (Uradni list RS, št. 16/2007 – ZASP–UPB3, 68/2008) dovoljujem / ne dovoljujem (ustrezno obkrožite), da se zgoraj navedena doktorska disertacija objavi v repozitoriju Dissertations and Thesis (Proquest) in dLib.si (NUK).

Podpis mentorja:

Podpis odgovorne osebe naročnika in žig:

Kraj in datum: Podpis avtorja/ice:

|  |
| --- |
| . |

Izjavljam, da je mentor seznanjen z indeksom podobnosti doktorske disertacije, ki je

Kraj in datum: Podpis avtorja/ice:

IZVLEČEK

…

**Ključne besede:**

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

**ARRS klasifikacija: 5.01.01, 2.07.08**

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

ABSTRACT

…

**Keywords:**

Education, tandem learning, machine learning, data mining.

**Math. Subj. Class. (2020):** 97D40, 97D60, 62P99

**ARRS classification: 5.01.01, 2.07.08**

Contents

[Uvod 10](#_Toc155434857)

[Teoretični del 10](#_Toc155434858)

[Pouk in učne oblike 10](#_Toc155434859)

[Kratko o pouku 10](#_Toc155434860)

[Neposredna in posredna učna oblika 10](#_Toc155434861)

[Delo v tandemu 10](#_Toc155434862)

[Začetki, razvoj in možnosti dela v tandemu 10](#_Toc155434863)

[Potek in struktura dela v tandemu 10](#_Toc155434864)

[Prednosti in slabosti dela v tandemu 10](#_Toc155434865)

[Strojno učenje in klasifikacija 10](#_Toc155434866)

[Uvod v umetno inteligenco 10](#_Toc155434867)

[Strojno učenje 10](#_Toc155434868)

[Naloga klasfikacije in izbire napovednih spremenljivk 11](#_Toc155434869)

[Naloga reduciranja dimenzije prostore 17](#_Toc155434870)

[Empirični del 18](#_Toc155434871)

[Raziskovalni problem, namen in cilji 18](#_Toc155434872)

[Raziskovalne hipoteze 19](#_Toc155434873)

[Metodologija 19](#_Toc155434874)

[Vzorec 19](#_Toc155434875)

[Zbiranje podatkov 19](#_Toc155434876)

[Obdelava podatkov 20](#_Toc155434877)

[Rezultati in interpretacija 20](#_Toc155434878)

[Deskriptivna statistika 20](#_Toc155434879)

[Test zanesljivosti 23](#_Toc155434880)

[Test normalnosti 23](#_Toc155434881)

[Test korelacije značilk 25](#_Toc155434882)

[Test pomembnosti značilk 26](#_Toc155434883)

[Test predikcije 28](#_Toc155434884)

[Test manjšanja dimenzije prostora oziroma vizualizacija podatkov 30](#_Toc155434885)

[Sklepne ugotovitve 31](#_Toc155434886)

[Literatura in viri 31](#_Toc155434887)

[Priloga A: Vprašalnik 37](#_Toc155434888)

Še kazalo preglednic in slik

# Uvod

300.000 do 600.000 znakov s presledki

# Teoretični del

## Pouk in učne oblike

### Kratko o pouku

…

### Neposredna in posredna učna oblika

...

## Delo v tandemu

### Začetki, razvoj in možnosti dela v tandemu

...

### Potek in struktura dela v tandemu

...

### Prednosti in slabosti dela v tandemu

...

### Spremenljivke, ki morebitno vplivajo na delo v tandemu

...

### Formiranje skupin

...

## 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 : 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).

#### Učni in testni podatki (dodaj literaturo...)

Začetni podatki so ključnega pomena za postavitev, učenje in testiranje algoritmov strojnega učenja. Navadno so zajeti v tabelo, kjer stolpci predstavljajo napovedne in ciljne spremenljivke. Ključen korak pri strojnem učenju je delitev začetnih podatkov na dve množici. Prvo imenujemo učna množica, na kateri se algoritem uči prepoznavati vzorce in relacije. Navadno vsebuje okoli 80 % začetnih podatkov. Več kot ima model podatkov, večja je navadno uspešnost algoritma. Druga množica se imenuje testna množica in služi za ocenjevanje točnosti ciljne spremenljivke. Na njej primerjamo rezultate algoritma z dejanskimi vrednostmi in tako vrednotimo točnost napovedi. Za oceno točnosti navadno uporabljamo metrike, ki jih opišemo v poglavju x. Disjunknost množic je za evalvacijo algoritmov nujna predpostavka. Včasih testa množica vsebuje še množico validacije, ki služi sprotnemu optimiziranju algoritma.

### Naloga klasfikacije in izbire napovednih spremenljivk

#### Naloga klasifikacije

.../splošno o nalogi/

#### Naloga izbire napovednih spremenljivk

Pri tej nalogi želimo izbrati, katere napovedne spremenljivke imajo največjo moč. Izmed vseh izbranih želimo izluščiti tiste, ki pustijo najmanjšo napako, če jih izpustimo ven iz napovednega modela (Huang et al., 2014). Dodatne prednosti, ki jih ta metoda nudi so morebitno izboljšanje samega modela, vpogled v delovanje procesa, ki ga generirajo podatki in izboljšanje napovedi na ne-vidnih podatkih (Vergara & Estévez, 2014). Metode, ki to omogačajo tipično delimo na tri skupine: Zavijalci (wrappers), vdelane metode (embedded methods), in filter metode (filter methods) (Guyon & Elisseeff, 2003).

Zavijalci vključujejo vključitev induktivnega učnega algoritma kot del postopka ocenjevanja različnih podmnožic značilk (Kohavi & John, 1997). Ti postopki običajno ocenjujejo uspešnost na podlagi stopnje klasifikacije, dosežene na testnem nizu. Čeprav zavijalci dejansko lahko privedejo do močnih rezultatov generalizacije, imajo opazen pomanjkljivost obsežnih računskih zahtev, še posebej pri uporabi na visokodimenzionalnih naborov podatkov. Poleg tega so dovzetni za težave, kot so pretirano učenje in občutljivost za inicializacijo metode, kar lahko omeji njihovo praktičnost (Guyon & Elisseeff, 2003).

Vdelane metode se lotijo drugačnega pristopa z integracijo znanja o specifični strukturi razreda funkcij, ki jih uporablja določen učni stroj (Lal et al., 2006). Vdelani postopki so v primerjavi z zavijalci običajno manj računsko zahtevni. Vendar pa so še vedno precej počasnejši od filtrov in pogosto prepleteni s značilnostmi učnega stroja, kar pomeni, da so izbrane značilke odvisne od specifičnega uporabljenega algoritma (Guyon & Elisseeff, 2003).

Filtri delujejo na osnovi popolne neodvisnosti med učnim strojem in podatki ter uporabljajo metriko, ki je neodvisna od induktivnega učnega algoritma za ocenjevanje podmnožic značilk (Wlodzislaw et al., 2003). Filtri, za razliko od zavijalcev, kažejo stopnjo robustnosti proti preprileganju. Glede samega rezultata naloge pa se pogosto odrežejo slabše kot skupini metod zgoraj (Guyon & Elisseeff, 2003).

#### Modeli uporabljeni za klasifikacijo

#### Algoritmi strojnega učenja so programi, ki nabor podatkov spremenijo v model strojnega učenja. Ta model postane matematična predstavitev vzorcev, skritih v podatkih. Model lahko na podlagi priučenih matematičnih pravil podaja napovedi na neznanih podatkih (CITAT). V tem poglavju bomo opisali nekaj algoritmov strojnega učenja, ki so bili preizkušeni za klasifikaciji uspešnosti učne metode. Opis bo bolj deskriptvne narave, k vsem algoritmom pa se lahko sklicujemo na knjigo (CITAT). Na literaturo se ob opisu algoritmov navadno ne bomo sklicevali, saj gre za splošno uveljavljene modele, ki so enolično določeni.

##### Naivni Bayesov klasifikator

Naivni Bayes (NB) predpostavi, da so atributi pri danem razredu med seboj neodvisni, kar sicer v realnem svetu pogosto ni izpolnjeno (S. Chen et al., 2020). Metoda najprej izračuna verjetnost, da nek primer pripada določenemu razredu pri danih vrednostih značilk. Opazovan primer nato uvrstimo v razred z najvišjo verjetnostjo. Klasifikator se zaradi preprostosti pogosto uporablja v realnem svetu (Wickramasinghe & Kalutarage, 2021).

##### K-najbližji sosedi

k-najbližji sosedi (KNN) (angl. K-Nearest Neighbor) je metoda, ki učne primeri predstavi v prostoru glede na atribute s pomočjo evklidske metrike. Za nove primere izračuna razdalje do že naučenih primerov in na podlagi tega s pomočjo pogojne verjetnosti določi kateremu razredu nov primer pripada. Vizualizacijo algoritma prikazuje slika x.

A screenshot of a video game

Description automatically generated

##### Odločitvena drevesa

Odločitvena drevesa (DT) (angl. Decision Trees) so vsestranski algoritmi, ki temeljijo na vsti pravil oziroma pogojev “Če -> Potem”. Vizualizacijo procesa odločanja kaže slika x.

A black background with a black square

Description automatically generated with medium confidence

##### Gaussian mixture

Model Gaussove mešanice (GMM) (angl. Gaussian mixture model) je verjetnostni model, ki predpostavlja, da so vse učne množice ustvarjene iz mešanice končnega števila Gaussovo porazdeljenih spremenljivk z neznanimi podatki. Delovanje si lahko predstavljamo kot gručenje, ki vključuje informacije o kovariančni strukturi podatkov in centrih latentnih Gaussovih krivulj.

A screenshot of a computer screen

Description automatically generated

##### Linear Discriminant Analysis

Linearna diskriminantna analiza (LDA) (angl. Linear discriminant analysis) je metoda, ki išče linearne kombinacije značilk, ki bi uspešno ločile razrede. Vizualizacija algoritma je enaka kot pri algoritmu SVM (slika x), le da ne gre za hiperravnino, temveč za kombinacijo premic.

##### ADAboost

AdaBoost (angl. Adaptive Boosting), daljše povečanje prilagoditev je ansambelski algoritem, ki podatke klasificira s pomočjo kombinacije večih osnovnih in šibkejših algoritmov (na primer odločitvenih dreves). Na podlagi rezultatov iterativnih (prejšnjih) klasifikacij prilagaja nove učne primere, kot kaže slika x. Najprej dodeli enako težo vsakemu opazovanju. Ko je bilo prvo drevo ovrednoteno za natančnost, se prilagodijo uteži za različna opazovanja. Opazovanja, ki jih je bilo enostavno razvrstiti, imajo znižane uteži, opažanja, ki jih je bilo težko razvrstiti, pa so povečana. Drugo drevo se ustvari z uporabo teh prilagojenih uteži, s ciljem, da bodo napovedi drugega drevesa natančnejše od napovedi prvega drevesa.

A screen shot of a game

Description automatically generated

##### Gradient krepitev

Modeli krepitev gradienta (GBM) (angl. Gradient Boost Model) so znani po svoji visoki natančnosti in povečujejo splošne principe, ki se uporabljajo v AdaBoost. Primarna razlika med modelom GBM in AdaBoost je v tem, da GBM uporabljajo drugačno metodo za izračun, kateri učenci napačno identificirajo podatkovne točke. AdaBoost izračuna, kje je model slabši, tako da preuči podatkovne točke, ki so močno ponderirane. Medtem GBM uporabljajo gradiente za določanje natančnosti učencev, pri čemer za model uporabljajo funkcijo izgube. Funkcije izgube so način za merjenje natančnosti prileganja modela naboru podatkov, izračun napake in optimizacijo modela za zmanjšanje te napake. GBM omogočajo uporabniku, da optimizira določeno funkcijo izgube na podlagi želenega cilja.

A group of red and blue dots

Description automatically generated

##### Metoda podpornih vektorjev

Metoda podpornih vektorjev (SVM) (angl. Support Vector Machine) predpostavi, da je večdimenzionalni učni prostor primerov linearno rešljiv. Učne primere nato ločimo s hiperravnino (na sliki x je hiperravnina premica, saj imamo le dva razreda). Ravnina, ki jo iščemo je najbolj oddaljena od vseh razredov.

A screenshot of a video game

Description automatically generated

##### Naključni gozd

Naključni gozd (RF) (angl. Random forest) je metoda, ki združi več odločitvenih ali regresijskih dreves. Prvi je opisal in predlagal metodo (Breiman, 2001).Vsako drevo je zgrajeno na nekoliko drugačni učni množici značilk. Končna rešitev je tista, za katero glasuje največ zgrajenih dreves. Vizualizacijo algoritma prikazuje slika x.

A black background with a black square

Description automatically generated with medium confidence

#### Evalvacija modelov strojnega učenja

Omejimo se na metrike, ki se uporabljajo za tri-stopenjsko klasifikacijo, saj se empirični del sklicuje le na ta tip strojnega učenja. Sklicevali se bomo na (Grandini et al., 2020) pri metrikah in na (Varoquaux & Colliot, 2023) pri križni validaciji.

##### Matrika zmede

Matrika zmede (angl. Confusion Matrix) predstavlja tabelo, katere vrednosti prikazujejo število primerov, ki so pravilno ali napačno napovedani za posamezni razred. Primer prikazuje slika x. Pravilna rezultata sta resnično pozitivno (TP) (angl. true positive) in resnično negativno (TN) (angl. true negative). Napačna rezultata sta lažno negativno (FN) (angl. false negative) in lažno pozitivno (FP) (angl. false positive). V 3-stopenjski klasifikaciji pravilne napovedi predstavljajo vse diagonalne vrednosti, vsi ostali pa negativne napovedi (Vujovic, 2021). V tem primeru imamo matriko velikosti 3x3, kot prikazuje slika x.

A screenshot of a game

Description automatically generated

##### Natančnost

Natančnost (angl. accuracy) predstavlja verjetnost, da model točno napove izid. Osnoven element metrike je posamezno opazovanje in vsako to opazovanje z enako težo vpliva na vrednost metrike. Tako moramo biti pozorni na število elementov v posameznem razredu. Tisti z večjim številom opazovanj bodo imeli večjo težo, torej je natančnost primerna takrat, ko nas zanimajo posamezni primeri in ne celoten razred. To se posebno opazi pri neuravnovešenih naborih podatkov (angl. imbalanced dataset), kjer je večina opažanj v enem razredu. S to metriko ni možno videti, v katerih razredih se algoritem obnese slabše. Po drugi strani pa je metrika izjemno intuitivna in lahka za razumeti, zato se v praksi tudi pogosto uporablja. Zavzema možne vrednosti med 0 in 1, kjer 0 predstavlja ničelno natančnost, 1 pa popolno natančnost (model pravilno napove vse testne primere).

Glede na oznake poglavja x, se natančnost izračuna po formuli

##### Preciznost in priklic

Preciznost in priklic (angl. precision in accuracy) sta metriki, ki sta uporabljeni predvsem v drugih metrikah.

These metrics will act as building blocks for Balanced Accuracy and F1-Score formulas. Starting from a two class confusion matrix: Figure 2: Two-class Confusion Matrix The Precision is the fraction of True Positive elements divided by the total number of positively predicted units (column sum of the predicted positives). In particular, True Positive are the elements that have been labelled as positive by the model and they are actually positive, while False Positive are the elements that have been labelled as positive by the model, but they are actually negative. Precision expresses the proportion of units our model says are Positive and they actually Positive. In other words, Precision tells us how much we can trust the model when it predicts an individual as Positive. The Recall is the fraction of True Positive elements divided by the total number of positively classified units (row sum of the actual positives). In particular False Negative are the elements that have been labelled as negative by the model, but they are actually positive. The Recall measures the model’s predictive accuracy for the positive class: intuitively, it measures the ability of the model to find all the Positive units in the dataset.

##### F1 score

F1-vrednost lahko razumemo kot uteženo povprečje preciznosti in priklica s pomočjo harmoničnega povprečja. Izračunamo jo po formuli

V tri-stopenjski klasifikaciji moramo vključiti preciznost in priklic vsakega razreda, ki nudi dva različna vpogleda v F1-vrednost.... Vključi [2008.05756.pdf, (arxiv.org)](https://arxiv.org/pdf/2008.05756.pdf,)

##### Križna validacija

Do zdaj smo vedno imeli v mislih učno in testno množico kot naključno izbrani. V strojnem učenju pa drugačno ločevanje podatkov porodi različne modele predikcije. Navadno vzamemo več različnih delitev množic z metodo križne validacije (angl. cross validation). k-križna validacija razdeli podatke v k množic, ki se imenujejo prepogibi (angl. folds) in so približno enake velikosti. V vsakem koraku se en prepogib uporabi za testiranje, ostali pa za učenje. Tako pokrijemo različne učne in testne podatke brez ponavljanja. Navadno vzamemo 10-20% veliko testno množico, kar pomeni, da vzamemo k = 5 ali k =10 prepogibov. Shemo 5-križne validacije prikazuje slika x.

A graph of red dots

Description automatically generated

##### Modeli, ki se navadno glede na zgornje metrike odzovejo bolje

...

#### Modeli izbrire napovednih spremenljivk

V cilju za določitev pomembnosti relacij med značilko in ciljno spremenljivko, se je pojavilo več numeričnih metod (Wei et al., 2015). Med njimi je ena najbolj uporabljenih Mutual information (MI), ki ima dve glavni prednosti. Sprejme lahko različne relacije, ki niso nujno linearne (Vergara & Estévez, 2014) in je robustna za spremenljivke, ki imajo veliko šuma (Li, 1990). Izračune MI, glede na tipe spremenljivk X in Y prikazuje tabela x.

|  |  |
| --- | --- |
| X in Y numerični |  |
| X in Y kategorični |  |
| X kategorična, Y numerična |  |

Dobljena vrednost je torej informacija, ki si jo delita obe slučajni spremenljivki. Kot smo opazili, jo lahko razrvstimo med filter metode.

Še RFE algoritem v povezavi z log? regresijo.

### 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). S tem načinom se znebimo morebitno zavajajočih in odvečnih značilk (Khalid et al., 2014).

Reduciranje dimenzije navadno poteka tako, da uporabljene značilke združimo v manj sintetičnih značilk, ki čim bolj ohranijo fundamentalno zgradbo originalov (Anowar et al., 2021). Povedano bolj poljudno, želimo manjšati število stolpcev, kjer se nova struktura stolpcev prepleta, oziroma »delanje krožnice iz sfere«. Obstaja tudi nogo načinov, kako metode reduciranja dimenzije klasificirati, ki jih lahko najdemo v (Anowar).

Oglejmo si delovanje enega najbolj znanih algoritmov te kategorije t-SNE, ki se najbolj uporablja v vizualizaciji. t-Distributed Stochastic Neighbor Embedding (t-SNE) je metoda, ki višje dimenzije projecira na nižje s tem, da ohranja tako lokalno kot globalno strukturo nabora podatkov (van der Maaten & Hinton, 2008). V disertaciji predstavimo delovanje algoritma poljudno, njegovo popolno vsebino s psevdokodo pa najdemo v (Melit Devassy & George, 2020) (ali damo v dodatek, toda to je huda statistika?).

1. **Izračun podobnosti**: Najprej t-SNE izračuna parno podobnost (s pomočjo Gaussove porazdelitve) med podatkovnimi točkami v visokodimenzionalnem prostoru.
2. **Inicializacija vložitev**: Naključno določimo začetne položaje vloženih točk v prostoru nižje dimenzionalnosti.
3. **Izračun podobnosti v nižji dimenziji**: Izračunamo podobnosti med vloženimi točkami v prostoru nižje dimenzionalnosti z uporabo podobne Gaussove porazdelitve.
4. **Optimizacija vložitev**: Iterativno prilagajamo položaje vloženih točk, da zmanjšamo razlike med podobnostmi v visokodimenzionalnem in nizkodimenzionalnem prostoru.
5. **Gradientni spust**: Za manjšanje skupnih razlik razdalj (Kullback-Leiblerjevih divergenc) je uporabljena metoda gradientnega spusta tako v visoki kot nizki dimenziji.
6. **Iteracija**: Postopek optimizacije se ponavlja, dokler se položaji podatkovnih točk v prostoru nižje dimenzionalnosti ne ustalijo, kar naj bi bilo stanje, kjer se razdalje (v pojmu bližine in ne eksplicitno razdalje) med posameznimi točkami v obeh prostorih najmanj razlikujejo.

# 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 skupno 89 dijakov s 54 rešenimi odgovori, ki so določali 14 spremenljivk. Osebe vključene v raziskavo so bili dijaki drugega (16 moških in 28 žensk) in tretjega (12 fantov in 33 žensk) letnika neke gimnazije v Ljubljani v šolskem letu 2023/24. Vzorec je bil neslučajnostni in namenski. Neuravnovešenost med spoloma je pričakovana in v skladu z vpisom na gimnazijo vključeno v raziskavo. Dostopa do socio-ekonomskega stanja (SES) nismo imeli.

### 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 Prilog.

### 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 je dostopen na (Bregant, 2023).

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

### Deskriptivna statistika

Deskriptivna statistika za vse obravnavane zvezne in ordinalne spremenljivke s kvantili je podana v tabeli x in y. Vizualni prikaz porazdelitve vseh spremenljivk je zbran na sliki x.

|  | Uspešnost | Ocena | Kvantitativna interakcija | Kvalitativna interakcija | Outperforming  partner | Razred | Profesor | Spol |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| mean | 2.4 | 3.4 | 2.2 | 2.1 | 2.1 | Categorical  (7 options) | Categorical  (4 options) | Categorical  (2 options) |
| std | 0.6 | 1.0 | 0.7 | 0.7 | 0.6 |
| min | 1.0 | 2.0 | 1.0 | 1.0 | 1.0 |
| 25% | 2.0 | 3.0 | 2.0 | 2.0 | 2.0 |
| 50% | 2.0 | 3.0 | 2.0 | 2.0 | 2.0 |
| 75% | 3.0 | 4.0 | 3.0 | 3.0 | 2.0 |
| max | 3.0 | 5.0 | 3.0 | 3.0 | 3.0 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Anksioznost | Motivacija | Introverzija | Intuicija | Čutenje | Presojanje |
| mean | 25.8 | 20.4 | 20.6 | 22.7 | 23.3 | 22.8 |
| std | 6.8 | 6.3 | 5.6 | 4.5 | 4.7 | 5.7 |
| min | 10.0 | 7.0 | 8.0 | 12.0 | 9.0 | 9.0 |
| 25% | 21.0 | 16.0 | 16.0 | 20.0 | 20.0 | 20.0 |
| 50% | 26.0 | 20.0 | 21.0 | 23.0 | 23.0 | 23.0 |
| 75% | 31.0 | 24.0 | 24.0 | 25.0 | 26.0 | 26.0 |
| max | 40.0 | 34.0 | 37.0 | 35.0 | 35.0 | 37.0 |

A group of graphs with different colored bars

Description automatically generated with medium confidence

Za boljši prikaz tako porazdelitve, kot kvantilov smo se za zvezne (psihološki tip osebnosti) poslužili violinskih škatel z brki, za kategorične pa tortnih diagramov.

A diagram of a diagram

Description automatically generated with medium confidence

A diagram of a chart

Description automatically generated with medium confidence

A green and blue pie chart

Description automatically generated

### 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. Za metriko smo uporabili Spearmanov koeficient, ki meri linearno zvezo med dvema normalno porazdeljenima spremenljivkama (De Winter et al., 2016; Rovetta, 2020; Schober et al., 2018).

A diagram of a person's personality

Description automatically generated with medium confidence

Figure 3: Zamenjaj za slovenska imena

V tabeli x imamo tri najbolj pozitivno in tri najbolj negativno korelirane spremenljivke. Vse imajo znatno korelacijo (0.3<|k|<0.5).

|  |  |  |
| --- | --- | --- |
| Par spremenljivk | | Spearmanov koeficient |
| Motivacija | Anksioznost | -0.50 |
| Spol (moški) | Čutenje | -0.36 |
| Spol (moški) | Motivacija | -0.31 |
| Spol (ženska) | Anksioznost | 0.34 |
| Interakcija kval. | Interakcija kvant. | 0.37 |
| Ocena | Motivacija | 0.43 |

### 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 (R.-C. 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 :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 : 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 3: 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

Bibal, A., Delchevalerie, V., & Frénay, B. (2023). DT-SNE: T-SNE discrete visualizations as decision tree structures. *Neurocomputing*, *529*, 101–112. https://doi.org/10.1016/j.neucom.2023.01.073

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

Bregant, B. (2023). *Tandem learning: Student dataset* (1.0) [dataset]. GitHub. https://github.com/borbregant/ai\_tandem\_learning

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

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

Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A novel selective naïve Bayes algorithm. *Knowledge-Based Systems*, *192*, 105361. https://doi.org/10.1016/j.knosys.2019.105361

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

Grandini, M., Bagli, E., & Visani, G. (2020). *Metrics for Multi-Class Classification: An Overview*. https://doi.org/10.48550/ARXIV.2008.05756

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

Harangi, B., Baran, A., & Hajdu, A. (2020). Assisted deep learning framework for multi-class skin lesion classification considering a binary classification support. *Biomedical Signal Processing and Control*, *62*, 102041. https://doi.org/10.1016/j.bspc.2020.102041

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

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

Huang, M.-L., Hung, Y.-H., Lee, W. M., Li, R. K., & Jiang, B.-R. (2014). SVM-RFE Based Feature Selection and Taguchi Parameters Optimization for Multiclass SVM Classifier. *The Scientific World Journal*, *2014*, 1–10. https://doi.org/10.1155/2014/795624

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

Li, W. (1990). Mutual information functions versus correlation functions. *Journal of Statistical Physics*, *60*(5), 823–837. https://doi.org/10.1007/BF01025996

Melit Devassy, B., & George, S. (2020). Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE. *Forensic Science International*, *311*, 110194. https://doi.org/10.1016/j.forsciint.2020.110194

*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

van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, *9*(nov), 2579–2605.

Varoquaux, G., & Colliot, O. (2023). Evaluating Machine Learning Models and Their Diagnostic Value. In O. Colliot (Ed.), *Machine Learning for Brain Disorders* (Vol. 197, pp. 601–630). Springer US. https://doi.org/10.1007/978-1-0716-3195-9\_20

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

Vujovic, Ž. Ð. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*, *12*(6). https://doi.org/10.14569/IJACSA.2021.0120670

Wei, M., Chow, T. W. S., & Chan, R. H. M. (2015). Heterogeneous feature subset selection using mutual information-based feature transformation. *Neurocomputing*, *168*, 706–718. https://doi.org/10.1016/j.neucom.2015.05.053

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

Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: Applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing*, *25*(3), 2277–2293. https://doi.org/10.1007/s00500-020-05297-6

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) |