# AI based influence of certain factors for teaching method selection: Using tandem learning in mathematics.

Bor Bregant orcid id 0009-0000-9331-5391

## Abstract

### Background

The main objective of highereducation institutions is to provide quality education to its students. One way to achieve this is is by introducing various teaching methods, one of which is tandem learning. Not everyone responds well to a one-size-fits-all method, and therefore, uncovering insights for predictive model selection tailored to individual students or classrooms becomes imperative for teaching institutions. The knowledge is hidden among the educational data set and is extractable through data mining techniques. The primary objective of the study was to identify the key variables that significantly influence student performance in tandem learning using machine learning algorithms.

### Methods

A sample of N\_0 high school students and N\_1 predictor variables has been used. The outcome of interest was a three state variable indicating whether the student responded well to implementation of tandem learning into education environment or not. Study tested which predictor variables were most important using RFE, SVM, chi^2, ... and mutual information. Dimensionality reduction was also performed using PCA and t-SNE analysis.

### Results

Predstavitev dobljenih rezultatov (v cifrah).

### Keywords

Assessment, education, machine learning, tandem learning, data mining, teching methods

### Math subject classification, MSC2020

97D40, 97D60, 62P99

## Introduction

- Kateri je problem, ki ga želimo rešiti? - Zakaj je ta problem tako pomembno znati rešiti? - Kako so ta problem reševali pred to raziskavo? - Katere so težave prej uporabljenih metod? Ali so te metode objektivne? Ali so zanesljive? - Kaj se predlaga? - Zakaj ta metoda BI LAHKO bila boljša? - Kaj o tem trdi literatura?

## Theoretical framework

### Teaching methods and tandem learning

Critic of frontal teaching and new theoretical didactics, psychological, pedagogic, sociologic findings and positive experience in practical work have lead to the development of new indirect forms of education process (Blažič et al., 2003). Based on strong research literatures various forms of small-group learning are effective in promoting greater academic achievement, more favorable attitudes toward learning, and increased persistence through SMET courses and programs (Roschelle et al., 2010).

Tandem learning is a special learning approach, where two students make an experiment together, formulate a report, solve a problem etc (Tomić 2002 Spremljanje pouka, Ljubljana: preveri se enkrat ali je direkt citat ali povzeto). It is a simple approach from organizational standpoint, as pair members have more chance for activity than in frontal teaching and group teaching, however they are not alone as in individual teaching method (Blažič et al., 2003). Simple diagram in figure Figure 1 depicts main components of group-learning relationship.

A diagram of a group of people

Description automatically generated

Figure : Relationships among interaction components of group learning (Slavin et al., 2003).

Many pedagogs, psychologists, sociologists and didactics say, that an individual in modern society is a member of many groups, so it is important, that students develop necessary social skills already in school. Imlementing group learning achieves five important goals (Peklaj, 2001): Students learn about eachother, group identity develops, students support each other, they learn to respect differences in between group members and students develop teamwork characteristics. Group learning has its pros, as well as cons. Pros according to (Puklek, 2001): better student performance, developing mutual support and help, developing different skills (cognitive, emotional-motivation (čustveno motivacijske?), social skills and understanding one-self.) and efficiency (ekonomičnost). Cons according to Puklek: Group goal over individual, lack of experience leading to ressentiment of learning method, member focuses only on task given to him, less effective due to member differences and inequality regarding involved work. (Kubale, 2015) adds that group work is difficult to perform in classes with large ammount of students. (Slavin et al., 2003) identifies four considerable theoretical views on the achievement effects of cooperative learning, them being: motivationalist, social cohesion, cognitive-developmental and cognitive-elaboration, the latter two focusing on the interaction among groups of students. These four perspectives can be considered complementary.

Five basic elements of cooperative learning: (1) positive interdependence; (2) face-to-face promotive interaction; (3) individual accountability and personal responsibility; (4) frequent use of interpersonal and small group social skills; and (5) frequent, regular group processing of current functioning (Johnson et al., 1991) direkten citat

Hundreds of studies have been conducted with main objective being being to determine the effects of cooperative learning on student achievement. We must keep in mind that this learning method is not only theoretical and a debate of research; it is used at some level by millions on teachers (Slavin et al., 2003) (v viru tudi dejanska cifra 81% v ameriki uciteljev to na dnevni ravni...). Many studies, which can be found in (Johnson & Johnson, 2011; Slavin, 1996; Webb, 1991) have found positive effect of for cooperative learning. (iz roschelle – tu notri tudi s kaksnim faktorjem posamezne raziskave)

Cooperative learning can be argued that high achievers could be help back by having to explain material to their low-achieving counterparts. However it would be equally debatable that students who give lectures to their counterparts learn more than those who receive them. Most studies however found equal benefits for high, average and low achievers (Slavin et al., 2003). direkten citat

Spremenljivke, ki vplivajo na delo v skupini – zakaj smo izbrali v raziskavi

In the quest to predict the effects of tandem learning on student performance, an array of variables must be considered to provide a comprehensive understanding of this dynamic educational approach. Examining the general factors, such as gender, class, professor, and previous grade, sheds light on the contextual background and baseline performance of students. Previous grade may not significantly impact tandem learning outcomes, while gender and class could exert a somewhat influential role (Nunar, 2020). Beyond these demographic aspects, the psychological dimensions of extroversion (Ramsay et al., 2000) and personality type come into play, shaping the way students engage and interact within tandem learning environments. Myers-Briggs Type Indicator (MBTI), which has become very popular in research world measures cognitive style in four dimensions: extroversion-introversion, sensing-intuition, thinking-feeling and judging-perceiving (Ramsay et al., 2000) [tu je v viru še razlaga kaj so te dimenzije in kako bi lahko vplivale in da so dvomi zakaj je ta pristop dober in obratno...]. Within the realm of tandem learning itself, variables like the quality and quantity of student interactions, whether a student outperforms their partner, and the presence of a teacher at the tandem station all come into focus. Notably, Tomić's findings reveal that group size does not substantially affect student performance (Tomić, 2003??? Poglej ce res). However, Puklek's work underscores the significant impact of students' personalities and knowledge, emphasizing the positive role of competitiveness on student performance (Puklek, 2001). By synthesizing these diverse factors, we can develop a more holistic framework for predicting the effects of tandem learning on student performance and tailor educational strategies accordingly.

### Machine learning and classification

Data mining is the process of uncovering hidden patterns, relationships, or insights within vast datasets through techniques from statistics and database management (Baradwaj & Pal, 2012). It involves data preprocessing to prepare information for analysis and utilizes methods such as clustering and association rule mining. In contrast, machine learning, a subset of artificial intelligence, focuses on building predictive models by allowing computers to learn from data and make decisions or predictions. The sequences of steps identified in extracting knowledge from data is shown in Figure 2 below.

A black arrow pointing to a square

Description automatically generated

Figure : Knowledge discovery process

### Connecting tandem learning and machine learning

Let us briefly discuss where AI is used in education today. We will focus mainly on the use of AI to support learning (student and teacher facing AI). Technologies can be considered in terms of whether they are mainly student teaching (with primarily instructionist approach), student supporting (primarily constructivist approach) or teacher supporting (which primarily help teachers do tasks they already do but faster or with less effort) (Holmes et al., 2019).

Beyond its broader applications, machine learning has been harnessed to predict student performance with remarkable precision. Leveraging the power of data analytics and advanced algorithms, machine learning models have been applied to forecast student success, identify at-risk learners, and tailor educational interventions. This transformative application of machine learning is exemplified by research conducted by (Siemens & Gasevic, 2012), which introduced the concept of "learning analytics" and demonstrated its potential in predicting student outcomes. Through this introduction, we will delve into the specific ways in which machine learning is harnessed to predict student performance and its profound implications for the education sector. Some other examples of predicting student performance with different metrics and models can be found in (Abana, 2019; Bhusal, 2021; Cortez & Silva, 2008; Kotsiantis et al., 2004; Minaei-Bidgoli et al., 2003).

### Feature selection: theoretical background

The goal of feature selection is to select the smallest feature subset given a certain generalization error, or alternatively finding the best feature subset with k features, that yields the minimum generalization error. We are reducing data structure complexity in order to identify important feature variables as a set of new training instances (Huang et al., 2014). Additional objectives of feature selection are as follows: (i) improve the generalization performance with respect to the model built using the whole set of features, (ii) provide a more robust generalization and a faster response with unseen data, and (iii) achieve a better and simpler understanding of the process that generates the data.(Vergara & Estévez, 2014).

Feature selection methods are typically categorized into three primary groups: wrappers, embedded methods, and filter methods, as referenced in previous research. Wrappers involve the incorporation of an inductive learning algorithm as part of the process of evaluating different feature subsets. 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. 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. 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) (4 odstavki zgoraj so povzeti po teh dveh virih)

Feature extraction, conversely, entails reducing the dimensionality of data by consolidating correlated features into synthetic ones, while retaining the fundamental characteristics of the original features (Anowar et al., 2021). That comes with several benefits including – Improvement of MLAs’ performance through less misleading and redundant features. – Avoidance of overfitting through fewer features, and therefore lesser assumptions by the model, and simpler the model. – Less computing time and much less storage is required with lower data dimensions. – More ease of data visualization and interpretation (Khalid et al., 2014).

For our research, we will use two of the most representative and successful FEA’s: PCA and t-SNE. For feature selection, we will use ...

## Empirical work

## Methodology

In the present research, the causal non-experimental method of pedagogical research is applied.

## Sample

The sample was comprised of N students from 11th and 12th grade of a Slovenian Gymnasium (i.e., high school).

## Procedure

After obtaining students’ (or their parents’, if the students were minors) signed informed consent and the school principals’ approval, we collected and examined the success of tandem learning in regards to several variables. Success was measured in 3 states (good, neutral and bad) and in 3 different forms (regarding learning, diversification and overall). (V članku bomo verjetno izbrali le najbolj statistično uspešno). Independent variables were in general sense (gender, class, professor, previous grade) in psychological sense (MBTI variables: extroversion-introversion, sensing-intuition, thinking-feeling and judging-perceiving) and in regards to tandem learning (qualitytive interaction, quantitativy interaction, whether student performed more than their partner and teacher presence at tandem station). Data was anonymized using a coding scheme, such that anonymity and objectiveness were assured in every step of the research. The collected data were accessible only to the researcher.

Data was collected following after students included in research were involved in tandem learning environment during the course of approximately one week. A portion of the class period was devoted to normal classroom work, while some portion of the class period was devoted to working in tadem. Randomization was not taken into consideration. Students were assigned into pairs in regards to their partner at the two seat desk.

MBTI spremenljivke: We utilized variables obtained from the MBTI test, specifically the Open Extended Jungian Type Scales (OEJTS) as a cost-effective alternative. The OEJTS was designed as an open-source alternative to the widely recognized Myers-Briggs Type Indicator (MBTI). Data was gathered from (*Fastest Myers-Briggs Test*, n.d.), which is based on (*Myers-Briggs/Jung Test: Open Extended Jungian Type Scales*, n.d.) both of which being available for public use like this under creative commons. (tu notri je še opis kako so rezultati izračunani (za v disertacijo) in še nekaj literature)

## Data analysis

The gathered data was analysed using Python programming language, primarily using pandas and scikit-learn libraries (dodaj katera verzija scikita). Raw dataset can be found on Zenodo, while statistical code is openly accessible on [GitHub](https://github.com/borbregant/ai_tandem_learning).

In suma, we modified all categorical data into integer type in the form of tidy data. Firstly, we performed data scaling (of normal distributed variables), then performed feature extraction with methods below and lastly dimensionality reduction with ....

## Results

### Student sample

Statističen opis kakšen je bil vzorec (intervali zaupanja za npr. cilnje spremenljivke, …). MBTI vzamemo zvezne spremenljivke, ki normalizira podatke, ki bi sicer implicirali bipolarnost ljudi. (Ramsay).

A group of blue and white bars

Description automatically generated

### Variable importance

Testi za določitev pomembnih spremenljivk.

wrapper, embedded in filter metode (feature\_selection)... (isti vir prednosti/slabosti)

wrapper je indukcija npr. rfe

embedded ze v algoritmu npr svm

filter – neodvisno od ML npr hi kvadrat, mutual information

PCA ni za feature selection ampak samo redukcija dimenzije, enako t-sne

A white rectangular object with black text

Description automatically generated

A diagram of different colored squares

Description automatically generated

### Model performance

Rezultati izbranih modelov

Table : Related feature variables

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Description | Possible values | Variable type |
| Gender |  | 0-1 (Male, female) | A priori state |
| Class |  | 0-6 (7 present classes) | A priori state |
| Teacher |  | 0-3 (4 teachers) | A priori state |
| Previous grade |  | 1-5 | A priori state |
| Extroversion score | From online test:  (intercitat) | 0-1 (with 0.01 step) | Psychological background |
| Sensing / intuition | From online test:  (intercitat) | 0-1 (with 0.01 step) | Psychological background |
| Thinking / feeling | From online test:  (intercitat) | 0-1 (with 0.01 step) | Psychological background |
| Judging / perceiving | From online test:  (intercitat) | 0-1 (with 0.01 step) | Psychological background |
| Qualitative interaction |  | 1-3 (little communication – lot of communication) | Tandem learning |
| Quantitative interaction |  | 1-3 (work was not productive - ???) | Tandem learning |
| Outperforming partner |  | 1-3 (worked less – outperform) | Tandem learning |
| Teacher presence |  | 1-3 (present a little – present a lot) | Tandem learning |

Table : Head of dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Student number | Performance | Gender | Class | Teacher | Extroversion score | ... | Teacher presence |
| 0 |  | M | 3B |  | 86 |  |  |
| 1 |  | F | 3B |  | 76 |  |  |
| .. |  |  |  |  |  |  |  |

## Discussion

…

## Conclusions and limitations

Study does not include a prediction whether tandem learning is overall effective or not. It simply includes which variables impact student response. Optimal number of principal components was also not taken into account, as with our study it is not necessary.

Some authors have suggested that students should not be forced to use learning approaches that do not suit their cognitive style.

## Acknowledgements

We extend our heartfelt gratitude to the study participants and their dedicated professors for their invaluable contributions. Additionally, we deeply appreciate the insightful comments provided by our esteemed reviewers and editors.

## References

Abana, E. C. (2019). A Decision Tree Approach for Predicting Student Grades in Research Project using Weka. *International Journal of Advanced Computer Science and Applications*, *10*(7). https://doi.org/10.14569/IJACSA.2019.0100739

Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). *Computer Science Review*, *40*, 100378. https://doi.org/10.1016/j.cosrev.2021.100378

Baradwaj, B. K., & Pal, S. (2012). *Mining Educational Data to Analyze Students’ Performance* (arXiv:1201.3417). arXiv. https://doi.org/10.48550/arXiv.1201.3417

Bhusal, A. (2021). *Predicting Student’s Performance Through Data Mining*. https://doi.org/10.48550/ARXIV.2112.01247

Blažič, M., Ivanuš-Grmek, M., Kramar, M., & Strmčnik, F. (2003). *Didaktika: Visokošolski učbenik*. Visokošolsko središče, Inštitut za raziskovalno in razvojno delo.

Cortez, P., & Silva, A. (2008). *Using data mining to predict secondary school student performance*.

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

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

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

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

Johnson, D. W., & Johnson, R. T. (2011). *Learning together and alone: Cooperative, competitive, and individualistic learning* (5. ed. [Repr.]). Allyn and Bacon.

Johnson, D. W., Johnson, R. T., & Smith, K. A. (1991). *Cooperative learning: Increasing college faculty instructional productivity*. School of Education and Human Development, George Washington University.

Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. *2014 Science and Information Conference*, 372–378. https://doi.org/10.1109/SAI.2014.6918213

Kotsiantis, S., Pierrakeas, C., & Pintelas, P. (2004). Predicting students’ performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, *18*(5), 411–426. https://doi.org/10.1080/08839510490442058

Kubale, V. (2015). *Skupinska učna oblika* (2. dopolnjena izd). Samozal. V. Kubale ; Piko’s Printshop.

Minaei-Bidgoli, B., Kashy, D. A., Kortemeyer, G., & Punch, W. F. (2003). Predicting student performance: An application of data mining methods with an educational web-based system. *33rd Annual Frontiers in Education, 2003. FIE 2003.*, *1*, T2A\_13-T2A\_18. https://doi.org/10.1109/FIE.2003.1263284

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

Nunar, N. (2020). *Izzivi skupinskega dela učencev* [Master’s thesis, Univerza na Primorskem]. https://repozitorij.upr.si/IzpisGradiva.php?lang=slv&id=12851

Peklaj, C. (2001). *Sodelovalno učenje ali Kdaj več glav več ve* (1. izd., 1. natis). DZS.

Puklek, M. (2001). Skupinsko delo: Kako ga oceniti? *Didakta*, *11*(60/61), 47–51.

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

Roschelle, J., Rafanan, K., Bhanot, R., Estrella, G., Penuel, B., Nussbaum, M., & Claro, S. (2010). Scaffolding group explanation and feedback with handheld technology: Impact on students’ mathematics learning. *Educational Technology Research and Development*, *58*(4), 399–419. https://doi.org/10.1007/s11423-009-9142-9

Siemens, G., & Gasevic, D. (2012). Guest Editorial—Learning and Knowledge Analytics. *Educational Technology and Society*, *15*(1–2).

Slavin, R. E. (1996). Research on Cooperative Learning and Achievement: What We Know, What We Need to Know. *Contemporary Educational Psychology*, *21*(1), 43–69. https://doi.org/10.1006/ceps.1996.0004

Slavin, R. E., Hurley, E. A., & Chamberlain, A. (2003). Cooperative Learning and Achievement: Theory and Research. In I. B. Weiner (Ed.), *Handbook of Psychology* (1st ed., pp. 177–198). Wiley. https://doi.org/10.1002/0471264385.wei0709

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

Webb, N. M. (1991). Task-Related Verbal Interaction and Mathematics Learning in Small Groups. *Journal for Research in Mathematics Education*, *22*(5), 366. https://doi.org/10.2307/749186