

УНИВЕРЗИТЕТ У БЕОГРАДУ

Факултет организационих наука

НАСТАВНО-НАУЧНОМ ВЕЋУ

Предмет: Подобност теме и кандидата Катарине Велинов за израду докторске дисертације

Одлуком Наставно-научног већа Факултета организационих наука-Универзитета у Београду 05-01 бр. 3/79-11 од 11.09.2024. године именовани смо за чланове Комисије за преглед и одбрану приступног рада и оцену научне заснованости теме докторске дисертације кандидата **Катарине Велинов** под насловом:

„Развој модела кластеровања за анализу успеха студената”

На основу материјала приложеног уз Захтев кандидата, Комисија подноси следећи

РЕФЕРАТ

1. Подаци о кандидату

1.1. Биографски подаци

Катарина Велинов рођена 11.02.1987. године у Босилеграду, завршила је Гимназију “Бора Станковић” у Врању на математичком смеру 2006. године. Исте године уписује Факултет организационих наука, смер за Информационе системе и технологије, који је и завршила 2010. године са просеком 9.16. Након завршених основних студија, завршила је и мастер студије 2012. године на истом факултету и студијском програму, са просеком 9.57. Мастер рад на тему “Унапређење електронског образовања применом пословне интелигенције” је одбранила са оценом 10. Образовање наставља уписом докторских студија на Факултету организационих наука на смеру за Оптимизацију и аналитику, 2022. године. Положила је све испите са просечном оценом 9.89. Објавила је научни рад на 50. Интернационалном симпозијуму за операциона истраживања из области неуронских мрежа.

Тренутно ради у приватној софтверској компанији Mozart д.о.о са седиштем у Београду. Гради своју каријеру пуних 13 година на радном месту Сениор софтвер инжењер. Претходно је радила у приватној софтверској компанији ASW Инжењеринг д.о.о у Београду на позицији Руководилац тима за развој софтверских производа, који обухватају целокупно пословање малопродајних објеката. У наведеној фирми је изабрана за најбољег радника у 2021. години.

1.2. Стечено научноистраживачко искуство

У наставку ће бити приказане научноистраживачке активности кандидата. Оне обухватају радове објављене на конференцијама и у часописима, од године уписа докторских студија, активности у извођењу наставе и преглед положених испита.

Радови објављени у научним часописима

Velinov, K., Milošević, P., Dragović, I. (2023). SELF-ORGANIZING MAPS FOR CLUSTERING STUDENTS OF THE FACULTY OF ORGANIZATIONAL SCIENCES. In D. Stojković, D. Petrović & S. Dimić (Eds.), 50th International Symposium on Operational Research (SYM-OP-IS 2023) (pp. 475 - 482).

Следи списак положених предмета на докторским студијама са оценама и ЕСПБ бодовима:

Називи предмета	Оцена	ЕСПБ
Методологија научно-истраживачког рада	10 (десет)	10
Неуронске мреже и системи - одабрана поглавља	10 (десет)	10
Управљање ланцима снабдевања - одабрана поглавља	10 (десет)	10
Статистика у менаџменту	10 (десет)	10
Фази логика и системи - одабрана поглавља	10 (десет)	10
Рачунарска интелигенција - одабрана поглавља	9 (девет)	10
Мултиваријациона анализа - изабрана поглавља	10 (десет)	10
Машинско учење - одабрана поглавља	10 (десет)	10
Временске серије и фрактали - одабрана поглавља	10 (десет)	10

Кандидат је дана 20.09.2024. године успешно одбранио приступни рад за израду докторске дисертације под називом „Развој модела кластеровања за анализу успеха студената”.

1.3. Оцена подобности кандидата за рад на предложеној теми

Узимајући у обзир:

- резултате остварене током досадашњег образовања;
- резултате досадашњих истраживања;
- радно искуство у области информационих система;

закључује се да је Катарина Велинов у потпуности квалификован и припремљен да тему докторске дисертације самостално истражује и пружи научно-стручне доприносе у тој области.

2. Предмет и циљ истраживања

Анализа успешности студирања је од великог значаја за даље унапређење образовног процеса. Успех студената игра кључну улогу у образовним институцијама, јер се често користи као мерило перформанси институције. Рано предвиђање учинка студената може помоћи универзитетима да предузму превентивне мере, као што је планирање одговарајуће обуке како би се побољшала стопа успеха студената. Технике машинског учења се у последње време интензивно користе за предикцију успеха студената. Студија која је анализирала 4519 публикација од 2000. до 2019. године (Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C., 2022) показала је да академске заједнице имају све већу заинтересованост за коришћење вештачке интелигенције у образовне сврхе. Главне истраживачке теме укључују интелигентне системе подучавања за специјално образовање, обраду природног језика за учење језика, едукативне роботе за АИ образовање, откривање законитости у подацима за предвиђање успеха, системе препорука за персонализовано учење и неуронске мреже за евалуацију наставе (Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C., 2022).

Ефикасна примена метода машинског учења захтева низ одлука, укључујући дефинисање студентског успеха, избор релевантних атрибута, и одабир најприкладнијих метода машинског учења за специфичан проблем. Коришћење метода машинског учења може бити комплексно и изазовно за особе без техничке позадине. Иако постоји много специјализованих софтвера, процес је и даље сложен и укључује многе одлуке. Проблем је што нема универзалног алгорита за кластеровање који би одговарао свим потребама, јер различити корисници (ученици, наставници, администратори, родитељи и саветници) могу имати различите интерпретације истих података и различите интересе у вези са генерисањем партиција или кластера из истог скупа података. Алгоритам који задовољава потребе једне групе корисника можда неће задовољити потребе друге групе. Стога, развијање једног алгорита за груписање који би задовољио све захтеве није могућ (A. Dutt, M. A. Ismail and T. Herawan, 2017; Alyahyan, E., Düşteğör, D., 2020).

За сваки специфичан проблем може се наћи оптималан модел који даје најбоље резултате, али то не значи да ће тај модел бити једнако добар за друге проблеме. Уместо тражења једног универзално најбољег модела, потребно је развити широк спектар модела који ће бити ефикасни у различитим ситуацијама.

Проблем истраживања докторског рада је испитивање могућности примене различитих модела кластеровања на образовни процес (K-means, self-organizing maps, fuzzy C-means), идентификовање кључних атрибута и проналажење оптималног модела за кластеровање студената који најбоље одговара специфичним потребама.

Предмет докторског рада је развој и примена новог модела кластеровања за анализу успеха студената. Сви добијени закључци биће детаљно објашњени и графички интерпретирани. Нови модел ће бити примењен на проблеме кластеровања и одлучивања. Тиме би се показало да нови приступ, осим теоријске оправданости, карактерише транспарентност, као и интуитивност и једноставност при примени.

Циљ докторског рада је да се уведе нови модел кластеровања за анализу успеха студената. Предложени приступи биће примењени на проблеме кластеровања и одлучивања.

Почетна листа библиографских извора који ће се користити приликом израде докторске дисертације:

- A. Dutt, M. A. Ismail and T. Herawan, (2017). A Systematic Review on Educational Data Mining, in *IEEE Access*, vol. 5, pp. 15991-16005, doi: 10.1109/ACCESS.2017.2654247
- Ahmad, N. B., Alias, U. F., Mohamad, N., & Yusof, N. (2019). Principal component analysis and self-organizing map clustering for student browsing behaviour analysis. *Procedia Computer Science*, 163, 550-559.
- Al-Barrak, M.A., & Al-Razgan, M.S. (2016). Predicting students' final GPA using Decision Trees: A Case Study.
- Alqurashi, E. (2019). Predicting student satisfaction and perceived learning within online learning environments. *Distance Education*, 40(1), 133–148.
- Alyahyan, E., Düşteğör, D., (2020). Predicting academic success in higher education: literature review and best practices. *Int J Educ Technol High Educ* 17, 3. <https://doi.org/10.1186/s41239-020-0177-7>
- Bhagat, K.K., Wu, L.Y., & Chang, C. (2016). Development and Validation of the Perception of Students Towards Online Learning (POSTOL). *J. Educ. Technol. Soc.*, 19, 350-359.
- Bishop, Christopher M. (2006). Pattern recognition and machine learning. New York, Springer
- Bowers, A. J. (2019). Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. *Practical Assessment, Research, and Evaluation*, 15(1), 7.
- Bravo-Agapito, J., Romero, S.J., Pamplona, S. (2020). Early Prediction of Undergraduate Student's Academic Performance in Completely Online Learning.
- Capó, M., Martínez, A.P., & Lozano, J.A. (2018). An efficient K -means clustering algorithm for massive data. *ArXiv*, abs/1801.02949.
- Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions. *Educational Technology & Society*, 25(1), 28-47.
- Cortez, P., & Silva, A. M. G. (2008). Using Data Mining to Predict Secondary School Student Performance. In A. Brito, & J. Teixeira (Eds.), *Proceedings of 5th Annual Future Business Technology Conference*, Porto, 5-12.
- El Guabassi, I., Bousalem, Z., Marah, R., Qazdar, A. (2021). Forecasting Students' Academic Performance Using Different Regression Algorithms. In: Motahhir, S., Bossoufi, B. (eds) *Digital Technologies and Applications. ICDTA 2021. Lecture Notes in Networks and Systems*, vol 211. Springer, Cham. https://doi.org/10.1007/978-3-030-73882-2_21

- Guanin-Fajardo, Jorge Humberto, Javier Guaña-Moya, and Jorge Casillas. (2024). Predicting Academic Success of College Students Using Machine Learning Techniques, *Data* 9, no. 4: 60. <https://doi.org/10.3390/data9040060>
- Hagan, M., Demuth, H., Beale, M., & DeJesus O. (2014). *Neural Network Design* (2nd ed.). Martin Hagan.
- Hamsa H., Indiradevi S., Kizhakkethottam J.J. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm.
- Hodge, V.J., Austin, J. A (2004). Survey of Outlier Detection Methodologies. *Artif Intell Rev* **22**, 85–126. <https://doi.org/10.1007/s10462-004-4304-y>
- Islam, Ariful, and Sanzida Tasnim. (2021). An Analysis of Factors Influencing Academic Performance of Undergraduate Students: A Case Study of Rabindra University, Bangladesh (RUB). *Shanlax International Journal of Education*, vol. 9, no. 3, 2021, pp. 127-135.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for Clustering Data*, Prentice-Hall
- Javornik, Š., Klemenčič Mirazchiyski, E. (2023). Factors Contributing to School Effectiveness: A Systematic Literature Review. *Eur. J. Investig. Health Psychol. Educ.* **2023**, 13, 2095-2111. <https://doi.org/10.3390/ejihpe13100148>
- Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and information technologies*, 13(1), 61-72.
- Klir, G.J., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic - theory and applications*.
- Kocak, O. Goksu, I, Goktas, Y. (2021). The factors affecting academic achievement: a systematic review of meta analyses. *International Online Journal of Education and Teaching (IOJET)*, 8(1). 454-484.
- Kohonen, T. (2001). *Self-Organizing Maps*, Springer Berlin, Heidelberg
- 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>
- Li, Y., Gou, J., & Fan, Z. (2019). Educational data mining for students' performance based on fuzzy C-means clustering. *The Journal of Engineering*, 2019(11), 8245-8250.
- Liu, J., Peng, P. & Luo, L. (2020). The Relation Between Family Socioeconomic Status and Academic Achievement in China: A Meta-analysis. *Educ Psychol Rev* **32**, 49–76 . <https://doi.org/10.1007/s10648-019-09494-0>
- Maaten, L.V., & Hinton, G.E. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9, 2579-2605.
- Martins, M. P. G., Miguéis, V. L., Fonseca, D. S. B., & Alves, A. (2019). *A data mining approach for predicting academic success – A case study*, (pp. 45–56). Cham: Springer.

- Merceron, A., & Yacef, K. (2008). Interestingness measures for association rules in educational data. In *educational data mining 2008*.
- Michael, D., Kyriakides, L. (2023) Mediating effects of motivation and socioeconomic status on reading achievement: a secondary analysis of PISA 2018. *Large-scale Assess Educ* **11**, <https://doi.org/10.1186/s40536-023-00181-9>
- Minaei-Bidgoli B., D. A. Kashy, G. Kortemeyer and W. F. Punch. (2003). Predicting student performance: an application of data mining methods with an educational Web-based system, *33rd Annual Frontiers in Education, 2003. FIE 2003.*, Westminster, CO, USA, 2003, pp. T2A-13, doi: 10.1109/FIE.2003.1263284.
- Osmanbegovic, E., Suljic, M. (2012). Data Mining Approach for Predicting Student Performance, *Economic Review: Journal of Economics and Business*, ISSN 1512-8962, University of Tuzla, Faculty of Economics, Tuzla, Vol. 10, Iss. 1, pp. 3-12
- Oyelade, O. J., Oladipupo, O. O., & Obagbuwa, I. C. (2010). Application of k-means Clustering algorithm for prediction of Students Academic Performance. *arXiv preprint arXiv:1002.2425*.
- Parker, J.D., Hogan, M. J., Eastabrook, J. M., Oke, A., & Wood, L. M. (2006). Emotional intelligence and student retention: Predicting the successful transition from high school to university. *Personality and Individual differences*, *41*(7), 1329–1336.
- Pérez, B., Castellanos, C., & Correal, D. (2018). *Predicting student drop-out rates using data mining techniques: A case study*, (pp. 111–125). Cham: Springer.
- Ramaswami, M., & Bhaskaran, R. (2009). A study on feature selection techniques in educational data mining. *arXiv preprint arXiv:0912.3924*.
- Rastrollo-Guerrero, Juan L., Juan A. Gómez-Pulido, and Arturo Durán-Domínguez. (2020). Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review, *Applied Sciences* 10, no. 3: 1042. <https://doi.org/10.3390/app10031042>
- Richard-Eaglin, A. (2017). Predicting student success in nurse practitioner programs. *Journal of the American Association of Nurse Practitioners*, *29*(10), 600–605.
- Romero C. and Ventura S. (2010). Educational Data Mining: A Review of the State of the Art, in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 6, pp. 601-618, doi: 10.1109/TSMCC.2010.2053532.
- Sari, I. P., Al-Khowarizmi, A. K., & Batubara, I. H. (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering for Grouping Students' Abilities in Online Learning Process. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, *2*(1), 139-144.
- Shi, C., & Li, X. (2022). Research on clustering algorithm based on improved SOM neural network. *Computational Intelligence and Neuroscience*, 2022.
- Shovon, M. H. I., & Haque, M. (2012). An Approach of Improving Students Academic Performance by using k means clustering algorithm and Decision tree. *arXiv preprint arXiv:1211.6340*.

- Tadese, M., Yeshaneh, A. & Mulu, G.B. (2022). Determinants of good academic performance among university students in Ethiopia: a cross-sectional study. *BMC Med Educ* **22**, 395. <https://doi.org/10.1186/s12909-022-03461-0>
- Tompsett J, Knoester C (2023) Family socioeconomic status and college attendance: A consideration of individual-level and school-level pathways. *PLoS ONE* 18(4): e0284188, <https://doi.org/10.1371/journal.pone.0284188>
- Wang, Q., Liu, L., Liu, Z., Qiao, W., & Liu, F. (2022). Application of Fuzzy Clustering in Higher Education General Management Based on Internet Environment. *Mathematical Problems in Engineering*, 2022.
- Willems, J., Coertjens, L., Tambuyzer, B., & Donche, V. (2019). Identifying science students at risk in the first year of higher education: the incremental value of non-cognitive variables in predicting early academic achievement. *European Journal of Psychology of Education*, 34(4), 847–872.
- Xing, W. (2019). Exploring the influences of MOOC design features on student performance and persistence. *Distance Education*, 40(1), 98–113.
- Yadav, S.K., Bharadwaj, B., & Pal, S. (2012). Data Mining Applications: A comparative Study for Predicting Student's performance. *ArXiv, abs/1202.4815*.
- Zhao, Ying & Karypis, George. (2002). Evaluation of Hierarchical Clustering Algorithms for Document Datasets. International Conference on Information and Knowledge Management, Proceedings. 10.1145/584792.584877.

3. Полазне хипотезе

Основна хипотеза:

- Груписање студената се може унапредити увођењем новог модела кластеровања.

Помоћне хипотезе:

- Могуће је дефинисати оптимални модел за кластеровање студената који најбоље одговара специфичним потребама анализе студената на факултету;
- Кластеровање студената на основу академских перформанси може идентификовати групе са специфичним образовним потребама или ризиком од неуспеха.
 - о Ова хипотеза тестира да ли се студенти могу ефикасно груписати према сличности у академским перформансама, и да ли се на основу тих група могу препознати они којима је потребна додатна подршка.
- Примена алгоритама кластеровања може побољшати прецизност предвиђања академског успеха студената у односу на конвенционалне методе анализе података.
 - о Испитује се да ли алгоритми кластеровања могу пружити додатну вредност у анализи образовних података и да ли могу унапредити тачност у предвиђању успеха студената.

4. Научне методе истраживања

Основне методе које ће се користити за реализовање истраживања су:

- Прикупљање и систематизација расположиве релевантне литературе;
- Критичка анализа модела кластеровања, с посебном пажњом на моделе K-means, Fuzzy C-means i Self-organizing maps;
- Развој новог модела кластеровања за анализу успеха студената;
- Провера својства предложеног модела;
- Експериментална провера предложених решења и спровођење тестирања над расположивим подацима;
- Експериментална провера постављених хипотеза са циљем њиховог потврђивања или оповргавања.

5. Очекивани научни допринос

Резултати који ће проистећи из истраживања у овом раду ће садржати већи број доприноса, међу којима се издвајају следећи:

- Преглед и систематизација праваца у развоју модела кластеровања за анализу успеха студената.
- Преглед и анализа постојећих модела кластеровања.
- Увођење новог модела кластеровања.
- Примена предложеног модела на проблеме кластеровања и одлучивања.
- Друштвени допринос се огледа у могућности примене предложеног модела у различитим областима.

6. План истраживања и структура рада

Планиране су следеће активности у циљу израде предложене дисертације:

- Преглед и анализа расположиве литературе;
- Развој нових решења;
- Имплементација и евалуација предложених решења.

Оквирна структура предложене дисертације је следећа:

У првом поглављу биће изложен проблем, предмет и циљ истраживања, и дефинисане хипотезе. Затим ће бити дат опис структуре докторског рада по поглављима.

У другом поглављу биће наведени основни теоријски појмови и дефиниције који ће бити основ овог рада.

У трећем поглављу овог рада биће изложене основе анализе успеха студената. Затим ће пажња бити посвећена моделима кластеровања који се користе у образовању, K-means, self-organizing maps, fuzzy C-means, приказ њихових предности и мана, као и поређење ових метода. У наставку ће бити изложен критички осврт на постојеће моделе кластеровања. На крају овог поглавља биће анализирани проблеми у примени кластеровања на реалне проблеме у области образовања, досадашња истраживања на тему анализе студената и унапређења предложена у литератури.

У четвртом поглављу предложиће се нови модел кластеровања за анализу успеха студената.

У петом поглављу ће се изложити могућности примене предложених приступа на различите проблеме. Конкретно, нови модел ће бити примењен на проблем кластеровања студената Факултета организационих наука у Београду.

У последњем поглављу биће дат закључак, са посебним освртом на постављене хипотезе, као и преглед доприноса дисертације и могући правци даљег истраживања.

На крају рада биће наведена литература која је коришћена приликом израде дисертације.

Оквирни садржај докторске дисертације

1. Увод

1.1. Предмет и циљ истраживања

1.2. Полазне хипотезе

1.3. Структура дисертације

2. Основни теоријски појмови

2.1. Анализа успеха студената

2.2. Примена метода кластеровања у образовању

2.3. Унапређења предложена у литератури

3. Алгоритми кластеровања

3.1. K-means алгоритам кластеровања

3.2. Fuzzy C-means алгоритам кластеровања

3.3. Self-organizing maps алгоритам кластеровања

3.4. Поређење алгоритама кластеровања

4. Нови модел кластеровања за анализу успеха студената

4.1. Мотивација и идеја увођења новог модела

4.2. Развој модела кластеровања

5. Примена новог модела кластеровања за анализу успеха студената

5.1. Опис извора података

5.2. Примена новог модела кластеровања

6. Закључак

6.1. Остварени допринос

6.2. Могући даљи правци истраживања

7. Литература

7. Закључак и предлог

Из изложеног се може закључити да кандидат Катарина Велинов испуњава све услове предвиђене Законом о високом образовању за одобрење израде докторске дисертације под насловом „Развој модела кластеровања за анализу успеха студената”.

Катарина Велинов поседује неопходна знања из информационих система, машинског учења и образовања за успешан рад на докторској дисертацији.

Тема припада ужој научној области Машинског учења, мултидисциплинарна је и актуелна. Добијени резултати могу допринети унапређењу образовања и могу имати практичну примену у образовним установама.

На основу свега наведеног, комисија предлаже Наставно-научном већу да прихвати предложену тему и одобри израду пријављене докторске дисертације. За ментора докторске дисертације предлаже се др Борис Делибашић, редовни професор Факултета организационих наука, Универзитета у Београду.

Ментор:

др Борис Делибашић, редовни професор
Факултет организационих наука, Универзитет у Београду

Чланови комисије:

др Милија Сукновић, редовни професор
Факултет организационих наука, Универзитет у Београду

др Александар Ђоковић, ванредни професор
Факултет организационих наука, Универзитет у Београду

др Небојша Бојовић, редовни професор
Саобраћајни факултет, Универзитет у Београду

Београд, 30.09.2024.