

Supervisor's review of the diploma thesis

Name and surname of the review author: doc. RNDr. Iveta Mrázová, CSc.

Name and surname of the thesis author: Diellor Hoxhaj

The title of the thesis: Techniques Applicable to the Analysis of Educational Data

The text of the review:

Diellor Hoxhaj elaborated his diploma thesis on higher education (HE) data and their analysis using recent data mining techniques. Successful practices adopted in higher education might have an immense social, psychological, industrial, and economic impact on the entire country and its inhabitants. Yet, low graduation rates pose a severe cross-cultural problem, especially in STEM-related programs. Despite the abundance of publicly available HE data sources and vast ongoing research, the area still lacks a unified approach to analyzing HE data and adequate interpretation of the obtained results. In addition to the standard statistical methods, the newer data mining techniques might add to gaining insight into productive HE management strategies. In particular, the social network-based methods might elucidate compelling associations within the analyzed data.

Therefore, the thesis aimed to review data mining topics suitable to the analysis of selected types of educational data. The student had to discuss the main paradigms relevant to decision trees (ID3, C4.5, CART, and related ensemble methods – bagging, random forests, and boosting). Further, he had to provide an overview and mutual comparison of the techniques applicable to pre-processing, visualization, and clustering, such as feature selection, k-means, and LVQ-based clustering. Finally, the social network analysis routines had to be evaluated for their potential to detect significant new data characteristics, like, e.g., the centrality measures (betweenness, closeness, or PageRank), and to reveal mutually closely tied communities in societal data. The student had to focus on some of the above topics in more detail and propose a suitable strategy for analyzing real-world educational data and implementing the models. Evaluating the obtained results and gaining experience are considered an essential part of the thesis.

The thesis is written in English, is well-organized, contains few typos, and has a neat graphical format. In addition to surveying the selected data mining mechanisms, the text also debates the educational and political systems of the countries from which the examined data originates. The data analysis and the choice of a suitable data mining model are accompanied by many supporting graphs and charts. To test the performance of the proposed solutions, the author used his implementations, which are attached to the thesis. For the discussion, I have the following questions to consider:

- 1) In the US graduation rates analysis, state the optimum parameters actually utilized for the SOM network and discuss the robustness of the solution to the set parameters. Further, present the number (or ratio) of the institutions comprised by the individual clusters and discuss possible transfers of the institutions between the individual clusters.
- 2) For the British Parliament data, compare the analyzed HE-based social network with that one emerging after the last election.
- 3) For the built decision trees (Tables 2.102, 2.109, and 2.114) also provide their confidence intervals to assess the stability of the techniques.

Hence, the thesis's core consists of reviewing, analyzing, and implementing three data mining methods relevant to the chosen tasks. The performance of the discussed solutions was experimentally evaluated and interpreted satisfactorily. The submitted thesis thus fulfills its original objective. The student was able to delve deeply enough into the data mining discipline and demonstrated his capacity to independently solve relatively demanding tasks in the area. A detailed discussion of related issues, the chosen approach to HE data scraping, and their analysis demonstrate the author's good orientation within the field. He proved his competence in scrutinizing the gained experience and results objectively.

Recommendation for the defense:

Based on the above reasons, **I recommend admitting this diploma thesis for the defense and agree to award Diellor Hoxhaj the Master's degree after a successful defense.**

I suggest not considering the thesis for the annual award.

In Prague, 30. 8. 2024

Signature: