REVIEW

of

a dissertation for obtaining the educational and scientific degree of 'Doctor' in doctoral program Informatics, field 4.6 "Informatics and Computer Science"

Author of the dissertation: Emiliano Maksim Mankolli

Dissertation topic: Optimization Methods for Machine Learning Applications,

Reviewer: prof. Dr Kiril Ivanov Simov, IICT-BAS

By order № 319 /06.12.2023 of the Director of IICT-BAS I was appointed as a Member of the Scientific Jury for the defense of the dissertation of Emiliano Maksim Mankolli for obtaining the educational and scientific degree of 'Doctor'. The set of materials presented by the candidate includes all necessary documents relevant to the procedure.

The doctoral student provided five publications related to the thesis. They are published in proceedings of conferences. Four of them are co-authored and for one he is the single author. One of the publication is published in Communications in Computer and Information Science, vol 1316. Springer (SJR (0.188), Computer Science - Quartile Q3). The provided publications are enough to support the defence of the dissertation.

The dissertation abstract (38 pages) in English comprehensive and correctly reflects the structure and content of the disertation. The abstract is also translated in Bulgarian.

In the Bulgarian abstract five scientific and three application contributions are presented. In the English text six of them are presented. It can be said that the two added in the Bulgarian translation further expand the contributions of the dissertation in the right direction. Thus, the contributions correctly reflect the achivments of the dissertation. I thing it is possible to add as a new contribution: the creation of sets of data to solve the problems posed in the dissertation, since their creation of such data is complex enough and, moreover, they can be used in future developments both by the dissertation and by other researchers of similar problems .

The topic of the disertation „Optimization Methods for Machine Learning Applications" is one of the most challenging with respect to the application of machine learning and is well

represented in the dissertation from the point of view of the applied field – recruitment process, based on different types of documents describing the qualities, knowledge and skills of the candidates. In this task, the unified representation over which to apply various machine learning approaches for classification, ranking, and evaluation represents a heavy cognitive load for the human resource management professionals. Due to this fact, the results of the dissertation, which demonstrate the theoretical and practical benefit of the application of integrated systems, including elements of Artificial Intelligence - methods of natural language processing, vectorization of words and texts and various methods of machine learning.

The dissertation of Emiliano Maksim Mankolli (128 pages) is written in English language. It consists of an introduction, three chapters, a conclusion, a list of the contributions of the dissertation, the publications related to the dissertation, a declaration of originality of the results, and references. The bibliography includes 131 titles in English. They contain fundamental and contemporary theoretical works in area of Natural Language Processing, Machine Learning, as well as works in the fields of recruitment and human performance evaluation. A very pleasant impression is also made by the fact that there is a list of abbreviations at the beginning of the thesis, and conclusions are included at the end of each chapter, including a summary of the results, limitations of the study, and directions for future work.

The Introduction presents in details the problems to which recruiters are facing in their work and how Machine Learning and Natural Language Processing methods can revolutionize the field. In this way, the doctoral student fully motivates the necessity and importance of the research presented in the dissertation and the results obtained. In addition to the preparation of the information contained in the applicants' documents, the reduction of the personal factor of human resource management specialists in the evaluations and decisions they make on the basis of this information.

The First chapter „Survey of the State of the Art: Machine Learning, Natural Language Processing, and Optimization" provides a very detailed introduction to the development, methods, and applications of the technologies used in the dissertation: Natural Language Processing; Machine Learning and Optimization Problems in Machine Learning. Natural Language Processing is described at several interrelated layers: tasks to be solved; methods and processing used and applications in the field of recruitment. The overview is complete providing

information about methods and tasks that are not used in the research reported within the dissertation. This shows a good level of knowledge of the doctoral student on these topics. The field of Machine Learning is also presented in a comprehensive way, covering all important approaches in the field. In addition to the two main approaches - Supervised Learning and Unsupervised Learning, methods such as Reinforcement Learning, Semi-Supervised, Active Learning Techniques, Deep Learning and Neural Networks, Transfer Learning and Ensemble Methods are also presented. These methods play an important role in Natural Language Processing. In the area of creation of Large Language Models, which are currently seen as the basis for the advancement of Artificial Intelligence. Models related to Machine Learning and Optimization Problems are also discussed in details.

At the end of the chapter a list of the objectives of the dissertation and the tasks to be solved is given.

The Second chapter „Streamlining the Candidate Pool: Optimization of Accuracy and Efficiency for Job Title Similarity" offers several solutions to the problem of screening large numbers of applicants. The task is presented as a problem of searching for similar positions within a set of job titles (80,000) and industry descriptions (700). The doctoral student gives a motivation why this issue is important in the field of recruitment. The proposed solution uses a combination of vectorization using the system word2vec (which is correctly described as a static approach to word embedding) and a combination of SVM (Support Vector Machines) and k-NN (k-Nearest Neighbors) for finding the similar position desptions. This approaches are described in detail in the dissertation. The procedure is as follows: word2vec is used to vectorize the job and industry descriptions, which are used to create the label data that is used by SVM and k-NN to solve the problem. The proposed approach is tested on the basis of the collected data. One question that is open is how to obtain the vector representation of the textual description, having in mind that word2vec only works with single word. Maybe the doctoral student used a approach based on word2vec modified for representation of whole text which is given in the references.

The second approach to the this task presented in the second chapter uses another word and text vectorization approach – BERT. BERT itself is well described in the dissertation. This second approach to finding the It also uses XGBoost (Extreme Gradient Boosting) to improve model performance. The application of these approaches is similar to that described in the first part of the

3

chapter. BERT is used to associate job descriptions with labels based on similarity to industry descriptions. XGBoost is then used to find similarity job descriptions. Then they are filtered by direct similarity calculation. An example also given in the chapter.

The Third chapter „Predictive Models for Job Success. Optimization of candidate selection strategy" describes a approach to pre-estimate the candidate success in the long term if she is hired for a given position. The chapter first describes the theoretical foundations for this assessment. The data used in the research presented in the chapter are descriptions of the applicant's professional development: professional experience, skills, previous jobs and others. These descriptions are represented as a feature set for the candidate, either by normalization of the numerical input or vectorization using BERT. After the feature sets are created, they are used to classify candidates as successful or not using XGBoost. The advantage of this approach over previous approaches is demonstrated.

The Conclusion of the thesis summarizes the work, the results that support the goals presented at the beginning of the thesis. The obtained results are shown to optimize the processing of applicant information and its efficient comparison with the relevant job information.

The dissertation has clearly expressed scientific and application contributions. The contributions are very well presented in the developed algorithms and architectures for working with job, industries and candidates descriptions. They also demonstrate the optimization of efficiency of solutions of these tasks from the field of recruitment. Thus, the objectives of the dissertation have been fully achieved.

Based on everything that has been said above regarding the merits of the thesis, I would like to state that I confidently propose to the esteemed Jury to award Emiliano Maxime Mancolli the scientific and educational degree of "Doctor".

19.02.2024 г.

Reviewer: *На основание ЗЗЛД*