deeper knowledge by taking advanced elective or concurrent courses.

There are a few very interesting observations from Table 2 of the case study. These observations can also offer some guidance on how to better embed analytics topics in an existing IT curriculum. First, an early embedding could be further strengthened in most programs. As can be seen, the analytics concepts are only sparsely covered at the 100-200 level courses, which implies that most students need to wait until their junior year to get exposed to this key area. Second, text data mining, as a whole knowledge area, should play a much stronger role in the IT curriculum given that most of the data in the modern world are in some unstructured format. Text is one of the most important unstructured data, in the forms of web pages, social media, emails, notes, and many more. The ability to model, store, access, and analyze large-scale text has become one of the most essential skills for future IT professionals. Third, analytics should be more comprehensively integrated into the human centered computing program to better support various human activities. The HCC area is largely empty in the current embedding map, which indicates the lack of support in this critical area for training next generation professionals in human computer interaction. Finally, some specific topics of significant importance are not adequately covered in the current programs. For example, time series analysis should carry a much higher weight in modern data analytics as many types of data, such as network traffic, stock exchanges, sensor readings, videos, speeches, among many others, are usually collected in real-time and fast changing. Students should be equipped with the necessary analytics skills to handle dynamic and fast changing data.

In addition to the above observations and guidance, different embedding types in Table 3 as discussed in our case study also suggest proper ways to incorporate analytics topics in an IT curriculum, especially into some existing courses. Among the four embedding types, LBE offers a convenient and light way to embed analytics topics that build upon the existing topics in a course but make extensions and connections to analytics topics. Thus, the change to an existing course will be minimum. For courses that cover general topics that are related to data, labs and case studies, are suitable choices given that analytics has become an essential way to deal with data. Finally, for courses that already focus on analytics, a semester-long project is usually necessary to equip students with sufficient skills to tackle real-world data analytics problems.

5. Conclusion

In this paper, we extend the Body of Knowledge for Data Science to formulate a formal BoK for Data Analytics, in recognizing the fast-increasing market demand in this key area. We propose to embed the key knowledge units from the four major knowledge areas in analytics into existing IT curricula. We further suggest different embedding types to cover these KUs, including LBE, lab, case study, and project, to ensure expected competency for the corresponding analytics skills of a student upon graduation. A case study is then provided to showcase how the relevant analytics KUs can be embedded into an existing IT undergraduate curriculum. Finally, some key insights and guidance are offered to further improve the embedded IT curriculum in order to meet the new requirements raised by the modern analytics pipeline.

REFERENCES

[1] Canada's Association of Information Technology (IT) Professionals, "A Guide to the Common Body of Knowledge for Computing and IT (CBOK)", Canadian Information Processing Society, Canada retrieved May 24, 2019 from http://www.cips.ca/CBOK.

[2] Computer Science vs. Data Science. Retrieved on 6/6/20 from https://www.onlineengineeringprograms.com/faq/computer-vs-data-science

[3] Data Science (n.d.). In Wikipedia. Retrieved June 1, 2020, from https://en.wikipedia.org/wiki/Data_science.

[4] I. B. Hassan and J. Liu, "Embedding Data Science into Computer Science Education," 2019 IEEE International Conference on Electro Information Technology (EIT), Brookings, SD, USA, 2019, pp. 367-372, doi: 10.1109/EIT.2019.8833753.

[5] IT2017 Task Group. 2017. Curriculum Guidelines for Baccalaureate Degree Programs in Information Technology. DOI: http://it2017.acm.org.

[6] Jai W. Kang, Qi Yu, Edward P. Holden and Xumin Liu. 2019. Complementing Course Contents Between IT/CS: A Case Study on Database Courses. In *Proceedings of ACM Special Interest Group in IT Education Conference (SIGITE'19), October 3-5, 2019, Tacoma, Washington, USA.* ACM, New York, NY, *USA, 6 pages.* https://doi.org/10.1145/3349266.3351414.

[7] Jai W. Kang, Qi Yu, Edward P. Holden, and Tae H. Oh. 2016. Security Requirements Embedded in MS Programs in Information Sciences and Technologies. In Proceedings of the 17th Annual Conference on Information Technology Education (SIGITE '16). Association for Computing Machinery, New York, NY, USA, 77–82. DOI:https://doi.org/10.1145/2978192.2978235.

[8] M. Tanwar, R. Duggal and S. K. Khatri, "Unravelling unstructured data: A wealth of information in big data," 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, 2015, pp. 1-6, doi: 10.1109/ICRITO.2015.7359270.

[9] Malcolm Haynes, Joshua Groen, Eric Sturzinger, Danny Zhu, Justin Shafer, and Timothy McGee. 2019. Integrating Data Science into a General Education Information Technology Course: An Approach to Developing Data Savvy Undergraduates. In Proceedings of the 20th Annual SIG Conference on Information Technology Education (SIGITE '19). Association for Computing Machinery, New York, NY, USA, 183–188. DOI:https://doi.org/10.1145/3349266.3351417

[10] Michael E. Whitman and Herbert J. Mattord. 2004. Designing and teaching information security curriculum. In Proceedings of the 1st annual conference on Information security curriculum development (InfoSecCD '04). Association for Computing Machinery, New York, NY, USA, 1–7. DOI:https://doi.org/10.1145/1059524.1059526.

[11] World Economic Forum and Jeff Desjardins. 2019. How much data is generated each day? Retrieved on 6/6/20 from https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

[12] Y. Demchenko, T. Wiktorski, J. Cuadrado Gallego and S. Brewer, "EDISON Data Science Framework (EDSF) Extension to Address Transversal Skills Required by Emerging Industry 4.0 Transformation," 2019 15th International Conference on eScience (eScience), San Diego, CA, USA, 2019, pp. 553-559, doi: 10.1109/eScience.2019.00076.