The Necessity and Public Concerns about Privacy in Big Data and Data Mining in Education

Chandra Wijaya  
School of Computer Science  
BINUS UniversityTangerang Selatan, Indonesia  
chandra.wijaya006@binus.ac.id

Mulia Darma  
School of Computer Science  
BINUS UniversityTangerang Selatan, Indonesia  
mulia.darma@binus.ac.id

Nicholas Adidharma Sukardi  
School of Computer Science  
BINUS UniversityTangerang Selatan, Indonesia  
nicholas.sukardi@binus.ac.id

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*Abstract*)

Keywords—Big data, Data mining, Privacy, Education, Public Opinion

Introduction

Along with the advances in technology, Big data and Data mining has gained and received more attention due to the implementation and capability of the technology in many fields. Big data and Data mining provides a lot of methods and techniques that can be implemented and used to process data. With the use of Big data and Data mining in the field of education, analysis of data in a large scale can be produced in order to help improve said field. More specific topic that has been chosen for this research paper is how the method, techniques and implementations have a potential to produce concerns about privacy issues and the public opinion of said concerns that can surface. From the application of Big data and Data mining itself, there has been research that uses Data mining to predict the percentage of dropout cases along with the factors that may contribute to the dropout case [1]. For the matter of privacy of Data mining and Big data, in a paper by Reidenberg, J. R., & Schaub, F. (2018) [2], policies to help the concerns and issues of privacy in the usage of Big data and Data mining is explored. This research paper aims to explore the public opinion regarding the issues of privacy that occurs in the application/usage of Big data and Data mining. In this paper we will use the quantitative method of survey(s) to gain the opinions regarding the privacy concerns of using Big data and Data mining in the field of education.

# Literature Review

The use of Big data and Data mining have been common in a lot of fields, where Big data and Data mining can be used to process and extract hidden information from the data on a large scale. With many uses of Data mining and Big data where there is a lot of technique, methods, and algorithms, which leads to the concern of security and privacy.

Big data not only used for processing data, but also can be used to obtain a teaching style for educators as paper from Ali, I. (2020) [3] said. Data mining also can be used to obtain hidden information like the cause of a result as the paper from Abu-Oda, G. S., & El-Halees, A. M. (2015) [1] said. Where data mining can be applied to get a reason why a student gets drop out and to predict dropout by using Decision Tree (DT) and Naive Bayes (NB). A paper by Asril, E., Wiza, F., & Taslim (2016) said that this technique can be used to search a pattern to calculate the competence of a student [4]. [2] The use of Big data also causes privacy concerns like the InBloom case that have been discussed in paper from Reidenberg, J. R., & Schaub, F. (2018).

From the paper we have reviewed, we found the use of big data and Data mining bring us an advantage in education, not only bring an advantage but it will cause a problem and privacy concern. Because of this we are researching the public opinion regarding the usage of big data and data mining and has chosen to use the quantitative method of using survey(s)

# Methodology

The Methodology that we used and would like to introduce to the readers is the methodology of using the quantitative method of using survey(s). As the question of this paper is aimed at determining the general public's concerns about the usage of Big data and the public’s opinion regarding the necessity of the usage of Big data in education, we needed qualitative data and the method that best suits in collecting said data is by using survey(s).

The data is collected via surveys that our team has made filled with the questions to our paper’s subject, the questions our team designed are aimed at gathering answers in form of opinions, the form that these questions take vary between multiple-choice questions and short answers questions. The survey takes form on an online questionnaire. The sample size is currently at the time of writing at 60 responses.

* The questions in the questionnaire are as follows, the first three questions are the introductory questions aimed at getting the personal information of the respondent.
* The next question “Apakah anda berasal dari jurusan Computer Science” (“Do you major in Computer Science?”) is aimed at getting the information whether the respondent has background education in Computer Science and can be used to analyze the other answers.
* Next is the question of “Apakah anda mengetahui mengenai Big Data” (“Do you know about Big data?”) to factor in respondents that do not major in Computer Science but do know about Big data and also those who major in Computer Science but don't/haven’t learned about Big data.
* Next the question “Apakah menurut anda penggunaan Big Data berbahaya?” (“Do you think the usage of Big Data is dangerous?”). This is aimed at getting the opinion about the usage of Big data.
* Next the question “Berdasarkan jawaban anda sebelumnya kenapa?” (“Based on your previous answer, why?”). This question is to gather the reason for the previous questions' answer that made them think the usage of Big data is dangerous and vice versa.
* Next the question “Menurut anda keamanan data adalah hal penting?” (“Do you think that data security is important?”). This question seeks answers from the respondents in the form of their opinion about data security.
* Next the question “Jika tempat anda (universitas, sekolah, tempat kerja) menggunakan big data. Apakah akan menjadi suatu hal yang baik?” (“If your university, school or workplace implements big data, do you think it is good?”). This question asks about the opinion of the respondent on how they would think if big data were to be implemented in their spaces.
* Next the question “Berdasarkan jawaban anda sebelumnya kenapa?” (“Based on your previous answer, why?”).

Our method for analyzing the data collected from the respondents answers in our questionnaire is by first transforming the data from the Google Form into .csv and then into an SQL database with the help of the online file converter into SQL, SQLizer. Our analysis consists of Content Analysis where we try to grasp and uncover the meaning of words in the answers. This Content Analysis is with the consideration of the question “Do you major in Computer Science?”.

Other Tools that we used to help us in our analysis is by using Google Colab to use the pyplot and toPandas to help us visualize and analyse the data that we got into more visible and easier form to understand.

# Result

##### References

1. Abu-Oda, G. S., & El-Halees, A. M. (2015). Data mining in higher education: university student dropout case study. *International Journal of Data Mining & Knowledge Management Process*, *5*(1), 15.
2. Reidenberg, J. R., & Schaub, F. (2018). Achieving big data privacy in education. *Theory and Research in Education*, *16*(3), 263-279.
3. Ali, I. (2020). OPINI: PERAN DAN KONTRIBUSI BIG DATA DALAM PENDIDIKAN DAN PELATIHAN KEPUSTAKAWANAN. MADIKA: Media Informasi dan Komunikasi Diklat Kepustakawanan, 5(1), 8-15.
4. Asril, E., Wiza, F., & Taslim (2016). Penerapan Data Mining Untuk Menggali Informasi Tersembunyi Dalam Big Data Nilai Mata Kuliah. Jurnal Teknologi Informasi & Komunikasi Digital Zone, 7(2), 129-134.