International Journal of Research and Review
DOI: https://doi.org/10.52403/ijrr.20211140
Vol.8; Issue: 11; November 2021
Website: www.ijrrjournal.com
Case Study *E-ISSN: 2349-9788; P-ISSN: 2454-2237*

Study of Potential Classification of Lost Students in College Based on Information Extraction on Text-Based Social Media; Case Study of Panca Budi Pembangunan University

Eko Hariyanto[1], Sri Wahyuni[2], Supina Batubara[3]

[1,2,3]Faculty of Science and Technology, University of Panca Budi Pembangunan, Medan, Indonesia

Corresponding Author: Eko Hariyanto

ABSTRACT

The main problem studied in this study is the large number of lost students who harm universities because of the difficulty of monitoring or monitoring as a preventive measure. Therefore, this research becomes very important to be done so that college institutions can make efforts to detect early (classification) of students who potentially cannot complete their studies on time or students who will drop out (DO). Thus, PT institutions through related parties such as academic guidance lecturers, academic bureaus and others can do initial prevention by providing the best solution or solution to the problems faced by students. This research aims to determine the training data model consisting of academic and non-academic factors (including the results of extracting information from social media). Furthermore, this model is used as a basis for classifying students who have the potential to "graduate on time", "graduate not on time", and "DO". The method approach used is quantitative with text mining computational algorithms for the process of extracting knowledge / information from social media which is further used in data training, as well as data mining computational algorithms for the process of classification of potential completion of student studies. The mandatory external targeted in the first year is the publication of the international journal Scopus Q4 and in the second year is the publication of the international journal Scopus Q3. For additional external targets in the first and second years respectively are the publication of international journals indexed on reputable indexers, ISBN teaching books and copyrights. The level of technological readiness (TKT) in this study up to level 2 is the formulation of technological concepts and applications to classify the potential completion of student studies using data mining.

***Keywords*:** student lost, knowledge/information extraction, data classification, text mining, data mining

INTRODUCTION

Student lost is a term used to describe students who do not complete their studies. Student lost has a bad impact on Universities (PT) because it affects the value of accreditation and reduces the revenue of PT. Table 1.1 shows the difference in the number of students of Panca Budi Pembangunan University (Unpab) who entered with the number of students who graduated. The absence of a model or system that can detect students who have the potential to become student lost, becomes an obstacle for parties related to Unpab in conducting monitoring or supervision as a preventive effort.

Table 1: Number of students per year

Year	Number Of Students Coming In	Number Of Students Graduating
2010	1776 People	199 People
2011	1979 People	630 People
2012	2155 People	83 people
2013	1914 People	166 people
2014	3286 People	81 people
2015	3774 People	-

There are various factors that cause student loss that have been studied by previous researchers such as gender, age, place of residence, status, GPA, lecture hours, non-academic activities and others [1][2][3]. In addition to the factors that have been studied by researchers before, in this study the research team will use additional factors, namely student activity in cyberspace, especially social media as a variable that will be studied to find out the potential of student lost.

Today, social media is widely used by students because it can provide ease in the learning process as an effort to improve achievement [4][5], but on the other hand it can also have a bad impact on their learning achievements [6]. Therefore, this research becomes very important to do because pt institution can know (extraction) information implied from student activity on social media accurately in an effort to detect early (classification) of students who could potentially not complete their studies on time or students who will drop out (DO). Thus, PT institutions through related parties such as academic guidance lecturers, academic bureaus and others can do the initial prevention by providing the best solution or solution to the problems faced by students.

The purpose of this study is to determine the training data model consisting of academic and non-academic factors (including the results of extracting information from social media). Furthermore, this model is used as a basis for classifying students who have the potential to "graduate on time", "graduate not on time", and "DO".

Research is a follow-up to the research roadmap of LPPM Unpab and the research roadmap of researchers that supports the Pembangunan of ICT products to improve service effectiveness. For more details on the relevance of this research to the LPPM Unpab research roadmap and the researcher's roadmap will be explained in the section

LITERATURE REVIEW

Text Mining

Text mining is the process of extracting implicit (implied) knowledge from unsalized textual data. Text mining is part of the concept of data mining in finding patterns (information or knowledge). The difference between the two lies in the input data where the data entered for data mining is structured data while text mining is unstructured data such as documents, text citations and others[7][8].

Data Mining

Table Roadmap LPPM Unpab 2014 - 2033

Field \Topic Flagship	**2014-2018**	**2019-2023**	**2024-2028**	**2029-2033**
Technology Information and Communication	Identify problems ICT infrastructure Identify problems System/ Platform Open-based Source Identify problems Increased ICT Content Device Technology Identification ICT and ICT supporters Expert decision support systems and computerized technology Identification of social issues of humanities, economics, culture, welfare based on communication and computerized technology	Strengthening ICT Infrastructure, Tools and Content Strengthening System/ Platform Open-based Source Strengthening supporters of expert-based decisions and computerized technology Strengthening social models of humanities, economics, culture, welfare based on communication and computerized technology	Infrastructure device models, tools and ICT content System device model/ Platform Open-based Source Expert decision-supporting device models and computerized technology Model device social models of humanities, economics, culture, welfare based on communication technology and computerization	Pembangunan Infrastructure ICT Pembangunan System/ Platform Open-based Source Pembangunan of expert and technology-based decision support models Computerized Pembangunan of social model models of social models of humanities, economy, culture, welfare based on communication and computerized technology

Data mining is a process of finding repeatedly (iteratively) and intensively with the aim to extract knowledge / information in the form of patterns, relationships, changes, rules, formulas or models from data sets. In general, the main role of data mining is [9][10]: Estimation is calculating the approximate value of a new object. Examples of algorithms are Neural Network [11], Support Vector Machine [12], and others. Prediction, which is to predict the value of a data. Examples of algorithms are Linear Regression, Neural Network, Support Vector Machine, and others. Classification, i.e. labeling a new object based on prior knowledge. Examples of algorithms are Naïve Bayes [13], K-Nearest Neighbor [14], and others. Clustering, which is grouping objects based on the information / attributes of the object. Examples of algorithms are K-Means, Fuzzy C-Means, and others. Association, i.e. looking for the relationship of one attribute with another. Examples of algorithms are FP-Growth, A Priori, and others.

Statistical Analysis
Research Roadmap

Researcher Roadmap Table 2019 - 2033

Linearity Topic Lecturer	**Research Which Has Been Done**	**Research Basis 2019-2023**	**Research Applied 2024-2028**	**Research Pembangunan 2029-2033**
Extraction of information from structured data(data mining)and unstructured data (TextMining)	1. Method Analysis Simple Additive Weighting and Profile Matching In The Evaluation of Lecturer Performance (Case Study: Pe University Pembangunan of Panca Budi) (2015). Implementation Application Monitoring Maintenance Truck at PT. Serdang Hulu (2016). Design a Wake System Application Dianosa Expert Plant Diseases Soursop with Forward Chaining Method (2017). Design Helmet Prototype Air Quality Gauge (2017). Application of Methods Certainty Factor Expert System Of Internal Medicine Prognosis with Traditional Medicine Solutions (2018). "Panic Motion" Application Design Based on Real Time Video (2018).	1. Determination Classification of Potential Student Lost in College By Information Extraction On Social Media Text-Based (Study University case Pembangunan of Panca Budi) 2. Determination of Interest Classification Student Talent By Information Extraction On Social Media Image-Based (Study University case Pembangunan of Panca Budi)	Information Academic 2. Design Wake up System "Counseling Bot" For Prevention Student Lost And integration System Search Talent Interests	1. Pembangunan Classification System Student Potential Lost Dan Interest Search Bakan Based Information Extraction On Social Media Multimedia-based.

MATERIALS & METHODS

The approach used in this research is a quantitative approach where the process of processing data from the phenomenon studied will be processed systematically. Data processing uses text mining and data mining computational algorithms. The data used is student data sourced from the academic bureau of Panca Budi Pembangunan University in the last five years and divided into two groups of study programs, namely the Science study programs and group social studies programs. Data collection techniques use sampling techniques.

The flow of stages in this study is divided into two years (figure 1 and figure 2)

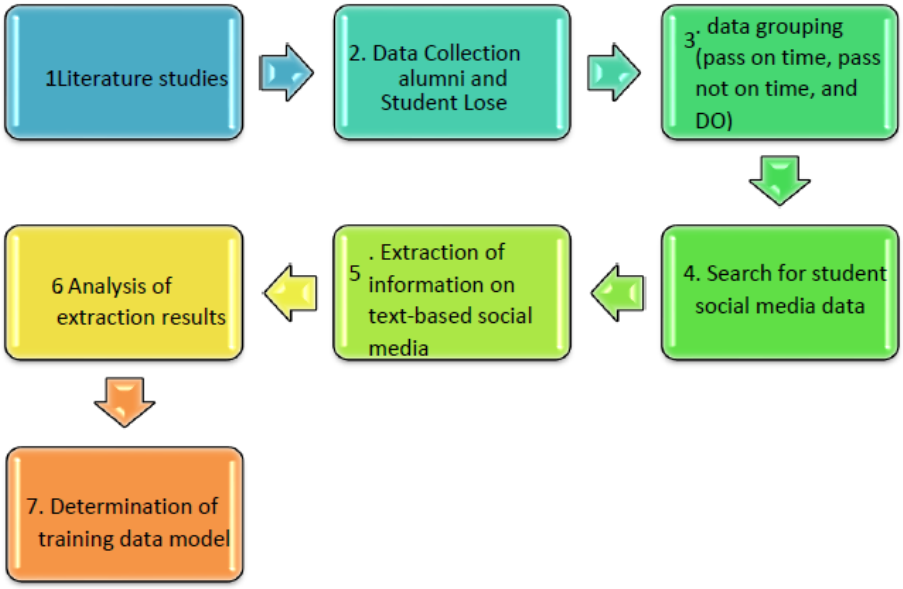

Figure 1. Year 1 research (2020)

Research Stage	Information
Literature studies	Collect and study literature from previous research
Collection of alumni data	Conduct hearings with academic bureaus and collect alumni data as well as student data that DO
Grouping data	Perform data filtering and grouping it into timely passes, untimely passes, and DO
Social media search	Collect social media links from DO alumni and students and browse social media
Extraction of information from social media	Perform a computational process to extract knowledge/ information from every social media searched
Analysis of extraction results	Analyze the similarities and differences of each result of knowledge / information obtained.
Determination of training data model	Determine the training data used for the process of classifying student lost potential from academic factors and non-academic factors

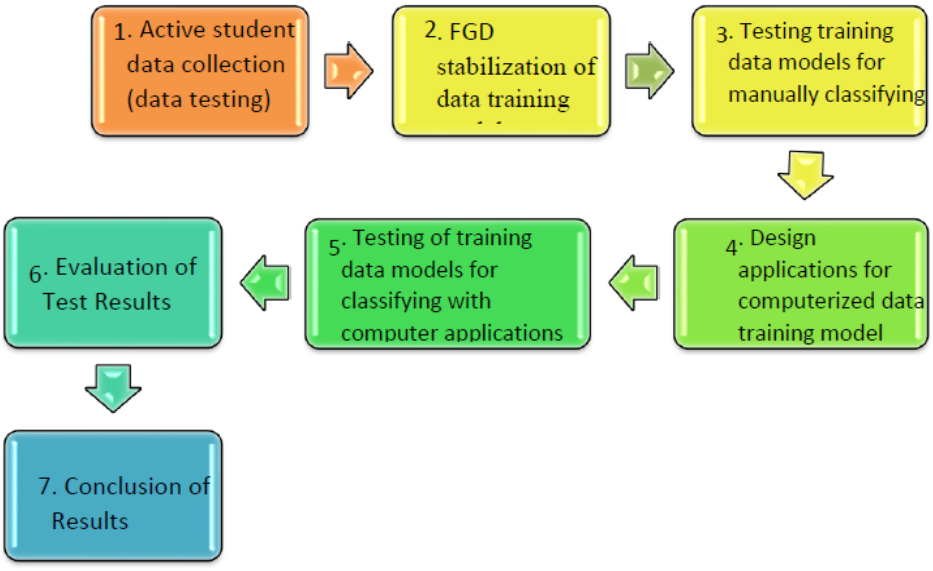

Figure 2. 2nd year of research (2021)

Research Stage	Information
Collection Student data active	Conduct hearings with academic bureaus and collect students who are still active who will be used as data testing
FGD stabilization of data training model	Focus Group Discussion (FGD) on training data models used as material to classify potential students completing studies
Manually testing the training data model	Conduct a manual data training model trial with several methods of classifying data(data mining)using the data testing above
Design training data test application	Design and build applications in accordance with the specifications of training data and classification methods used.

Table Continued…	
Computerized testing of training data models	Test data training models with applications created using data testing and some data classification methods (data mining).
Evaluation of Test Results	Analyze and evaluate the results of training data tests manually and computerized.
Conclusion of Results	Conclude the results of classification at the best conditions (best case) and in the worst conditions / conditions (worst case)

RESULT

Relevance of PT Roadmap and Research Roadmap

	Roadmap Lppm Unpab 2014-2033						**Relevance Of Roadmap Pt With Roadmap Researchers**
Field \Topic Flagship			**2019-2023**		**2024-2028**	**2029-2033**	
Technology Information and Communication		1. 2.	System Strengthening/ Open-Based Platform Source Strengthening supporters of expert and technology-based decisions Computerized	1. 2.	Device model System/ Platform Open-based Source Expert and technology-based decision support device models Computerized	Pembangunan System/ Platform Open-based Source Pembangunan of expert and technology-based decision support models Computerized	Roadmap researchers in accordance with the roadmap of universities that support the creation of ICT products to increase the effectiveness of public services
	Roadmap		Researchers		2019-2033		
Linearity Topic Lecturer	Research Thathas Been Done		Research Basis 2019-2023		Research Applied 2024-2028	Research Pembangunan 2029-2033	
Extraction of information from structured data (DataMining)and unstructured data (TextMining)	Analysis of Simple Additive Weighting and Profile Matching Methods in Performance Evaluation Lecturer (Case Study: Pembangunan Panca Budi University) (2015). Implementation of Monitoring Application Periodic Maintenance of Trucks at PT. Serdang Hulu (2016). Design To Build Expert System Application Dianosa Soursop Plant Disease with Forward Chaining Method (2017) Prototype Design of Air Quality Measuring Helmet (2017). Application of Certainty Methods Of Expert System Factors Diagnosed with Internal Medicine with Traditional Medicine Solutions (2018). "Panic Motion" Application Design Based on Real Time Video (2018).	1. 2.	Classification Determination Potential Student Lost in College Based on Extraction Information on the Media Text-Based Social (Study University case Pembangunan of Panca Budi) Determination of Interest Classification Student Talent Based on Extraction Information on the Media Image-Based Social (Study University case Pembangunan of Panca Budi)	1. 2.	Design a Wake Classification System Student Potential Lost and Integration With the System Information Academic Design a Wake System "Counseling Bot" For Prevention Student Lost Dan System Integration Interest Search Talent	1. Pembangunan Classification System Potential Student Lost And Search Bakan's Interests By Extraction of Information on Multimedia-Based Social Media.	

DISCUSSION

Year 1 Schedule Table (2020)

No.	Activity Name	Moon											
		1	2	3	4	5	6	7	8	9	10	11	12
1.	Literature studies												
2.	Collection of alumni and student data DO												
3.	Grouping data that passes on time, passes not on time, and DO												
4.	Search for student social media data												
5.	Text-based social media data processing (information extraction) with Text Mining												
6.	Progress report												
7.	Analysis of information extraction results												
8.	Preparation and publication of scientific articles												
9.	Determination of training data model												
10.	Final report drafting												
11.	Monitoring and evaluation												
12.	Log Book Fill												

2nd Year Schedule Table (2021)

No.	Activity Name	Moon											
		1	2	3	4	5	6	7	8	9	10	11	12
1.	Active student data collection (data testing)												
2.	FGD stabilization data training model												
3.	Testing training data models for manually classifying												
4.	Design applications for computerized data training model tests												
5.	Testing of training data models for classifying with computer applications												
6.	Preparation and publication of scientific articles												
7.	Progress report												
8.	Evaluation of test results												
9.	Conclusion of results												
10.	Final report drafting												
11.	Monitoring and evaluation												
12.	Log Book Fill												

CONCLUSION

This research becomes very important to do so that college institutions can make efforts to detect early (classification) of students who are potentially unable to complete their studies on time or students who will drop out (DO). Thus, PT institutions through related parties such as academic guidance lecturers, academic bureaus and others can do initial prevention by providing the best solution or solution to the problems faced by students. This research aims to determine the training data model consisting of academic and non-academic factors (including the results of extracting information from social media). Furthermore, this model is used as a basis for classifying students who have the potential to "graduate on time", "graduate not on time", and "DO". The method approach used is quantitative with text mining computational algorithms for the process of extracting knowledge / information from social media which is further used in data training, as well as data mining computational algorithms for the process of classification of potential completion of student studies.

Acknowledgement: None

Conflict of Interest: None

Source of Funding: None

REFERENCES

1. S. Wahyuni, K. S. Saragih and M. I. Perangin-angin, "Implementation of Decision Tree Method C4.5 To Analyze Dropout Students," Ethos: Journal of Research and Devotion (Science & Technology), vol. 6, no. 1, pp. 42-51, 2018.
2. A. P. U. Sembiring and M. Ginting, "Analysis of Factors Influencing Student Resignation With Data Mining Add-Ins Application – Case Study on Microskil

Stmik," SIFO Mikroskil Journal, vol. 6, no. 2, pp. 139-146, 2013.
3. F. Imran, B. Susetyo and A. H. Wigena, "Identification of Factors Related to College Dropouts at IPB Class of 2008 Using Survival Analysis," Xplore: Journal of Statistics, vol. 1, no. 2, pp. 1-6, 2013.
4. Rumyeni, "Social Media Acceptance Among Pekanbaru City Students," Journal of Communication Sciences, vol. 8, no. 2, pp. 117-132, 2017.
5. I. Mutia, P. Irfansyah and L. P. W. Adnyani, "The Influence of Facebook Social Networking on The Learning Achievement of Informatics Engineering Students at University," Journal of Informatics Education and Research, vol. 2, no. 2, pp. 136-141, 2016.
6. W. J. Drakel, M. H. Pratiknjo and T. Mulianti, "Student Behavior In Using Social Media At Sam Ratulangi Manado University," Holistic : Journal of Social and Cultural Anthropology, vol. XI, no. 21A, pp. 1-20, 2018.
7. T. Jo, Text Mining: Concepts, Implementation, and Big Data Challenge, Switzerland: Springer, 2018.
8. G. Miner, J. Elder IV, T. Hill, R. Nisbet, D. Delen and A. Fast, Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications, Waltham: Academic Press, 2012.
9. S. Pramana, B. Yuniarto, S. Mariyah, I. Santoso and R. Nooraeni, Data Mining With R Programming: Concept and Implementation, Bogor: In Media, 2018.
10. S. Tuffery, Data Mining and Statistics for Decision Making, Chichester: John Wiley & Sons, 2011.
11. A. Shabani, K. A. Ghaffary, A. R. Sepaskhah and A. A. Kamgar-Haghighi, "Using the artificial neural network to estimate leaf area," Journal Scientia Horticulturae, vol. 216, pp. 103-110, 2017.
12. W. Yao, C. Zhang, H. Hao, X. Wang and X. Li, "A support vector machine approach to estimate global solar radiationwith the influence of fog and haze," Renewable Energy Journal, vol. 128, no. A, pp. 155-162, 2018.
13. F. Harahap, A. Y. N. Harahap, E. Ekadiansyah, R. N. Sari, R. Adawiyah and C.B. Harahap, "Implementation of Naïve Bayes Classification Method for Predicting Purchase," in 6th International Conference on Cyber and IT Service Management (CITSM), Parapat, Indonesia, 2018.
14. G. A. Sandag, N. E. Tedry and S. Lolong, "Classification of Lower Back Pain Using K-Nearest Neighbor Algorithm," in 6th International Conference on Cyber and IT Service Management (CITSM), Parapat, Indonesia, 2018.

How to cite this article: Hariyanto E, Wahyuni S, Batubara S. Study of potential classification of lost students in college based on information extraction on text-based social media; case study of Panca Budi Pembangunan University. *International Journal of Research and Review*. 2021; 8(11): 325-331. DOI: *https://doi.org/10.52403/ijrr.20211140*

******