Walisongo Journal of Information Technology, Vol. 4 No. 2 (2022): 87-94
DOI: http://dx.doi.org/10.21580/wjit.2022.4.2.12676

Exploratory Data Analysis of Exact Science and Social Science Learning Content on Digital Platform

Mambang[1,*], Haniffah Sri Rinjani[1], Muhammad Zulfadhilah[1], Finki Dona Marleny[1], Septyan Eka Prastya[1], Subhan Panji Cipta[2]

[1]Department of Information Technology, Faculty of Science and Technology, Sari Mulia University, Banjarmasin, Indonesia
[2]Department of Informatics, Faculty of Engineering, University of Muhammadiyah Banjarmasin, Indonesia
[*]Corresponding author: mambang@unism.ac.id

Abstract

Data is one of the essential aspects in providing new information and new knowledge so that the data exploration process can provide policies on a decision for many sectors. Exploratory Data Analysis in this paper begins with collecting datasets contained on the Youtube digital platform. The dataset used was 30 samples found on the top page of youtube in each keyword. After conducting the Exploratory Data Analysis process, we found new learning content on the digital youtube platform. From the Exploratory Data Analysis that has been carried out, we also find different variations of the analysis's variables. The duration variable shows the result that the total duration of the overall duration in mathematics learning content that includes the exact field is less than the psychology learning content included in the non-exact field. Meanwhile, the overall number of views on mathematics learning content is more than the number of views on psychology learning content. From the collecting dataset that we have made, showing a considerable number of views is undoubtedly the key to equitable distribution of information and knowledge for all users. More innovation and creating learning content are expected to encourage increased human development.

Keywords: Explanatory data analysis, Exact and non-exact, Learning content, Digital platform, Python libraries

1 Introduction

Digital platforms are evolving along with many smart device users who continue to experience significant growth. Digital platforms such as Youtube that contain learning content can be used as one of the supporters of the learning process.

ISSN: 2715-0143 (online), 2714-9048 (print)
http://journal.walisongo.ac.id/index.php/jit/index

The combination of videos, comments, the number of users, and the number of users who like the content on digital platforms such as Youtube create interactions between content creators and users (Tanskanen, 2021). The ease of accessing learning content with a good internet connection infrastructure benefits learning content users. As one of the sources of information, youtube is very popular as a digital platform (Baran and Yilmaz Baran, 2021). Data is one of the essential aspects in providing new information and new knowledge so that the data exploration process can provide policies on a decision for many sectors. Exploratory Data Analysis includes the critical process of initial investigative tests on data to identify patterns, find anomalies, test hypotheses, and examine assumptions through summary statistics and graphical (visual) representations. The use of exploratory data analysis in many fields can provide new knowledge to the data carried out by the analysis so that the results of the analysis can be implemented (Sorg and Khobzi, 2022). Data exploration analysis can provide broad insights into many industry sectors and organizations (Hammouchi et al., 2019). Exploratory data analysis reveals the shape of a set of data and generates statistical distributions (Adeniyi et al., 2020).

Data literacy is critical in the current industrial era 4.0 because it can provide value in many fields by making data analysis and exploration. The industrial field and the organization are very dependent on data, so it is necessary to prepare quality human resources to manage data. Digital platforms that have experienced significant improvements need to be carried out Exploratory Data Analysis both on content and also users. The ease of accessing learning content through digital platforms needs to be optimized so that it can encourage the improvement of new competencies and insights for its users.

In comparing our paper, some previous studies discuss exploratory data analysis such as Whitelock-Wainwright et al. (2021) analyze students' expectations of learning services. The findings from their study show that students' expectations regarding the ethical and privacy variables of learning services are consistent across all groups. However, the expectations of the service's features vary considerably. Ahmadi et al. (2020) make a data analysis of ambidexterity that drives innovation in the manufacture of SMEs. This study shows exploitative dominant balanced ambidexterity as the optimal strategy composition in SMEs. Li et al. (2021) make a factor analysis of constructivist exploration in a survey of the learning environment of engineering graduates. This paper presents a questionnaire data collection on first-year engineering students' perceptions of constructivist practice in the learning environment. Of the 293 participants sampled by the questionnaire, 274 completed the questionnaire with a response rate of 93.515

Our paper focuses on data on the duration of learning content and the number of views on learning content. How we explore the data created is to do dataset analysis using several libraries contained in Python, such as Matplotlib, Pandas, Numpy, and also Sklearn. What is the purpose of our paper can be shown in the process by exploring data on learning content contained on Youtube using samples of mathematics and psychology

learning content. The results of this data exploration help find information from both samples of variables, thus providing new information in the context of learning content contained on digital platforms such as Youtube.

2 Material and Method

The collection of datasets is obtained by typing keywords on the Youtube search menu, such as learning mathematics and learning psychology. Mathematics learning content (BM) enters the same field, and learning psychology (BP) enters the non-exact field. The dataset used was 30 samples found on the top page of youtube in each keyword. In the table below, only ten examples of datasets are shown that are used in this Exploratory Data Analysis. The analysis is performed to identify the structure and interrelationship of the data (Ide et al., 2020).

2.1 Python Library

Compared to other programming languages, Python is more popular because it is very productive and has many types of capable libraries, such as TensorFlow, NumPy, SciPy, Pandas, and Matplotlib, Keras, SciKit-Learn, PyTorch, and Scrapy. A python library is a collection of related modules containing code that can be used repeatedly in different programs. The existence of libraries makes Python programming more superficial and more convenient for programmers because there is no need to write the same code repeatedly for different programs. Python libraries that are available open-source provide convenience in carrying out data analysis and testing (Zanovello et al., 2022). The python library can be run on many operating systems (Chacon-Hurtado and Scholten, 2021). Python libraries have also provided various types of algorithms that can be used according to the characteristics of the data used to analyze and test datasets (Meyer, 2021). Python allows users to create interface design graphics (Brandstetter et al., 2021).

2.2 Proposed Method

Exploratory Data on learning content carried out in this paper is carried out by the analysis method. Data analysis is a data processing process to find helpful information that can be used as a basis for decision-making for the solution of a problem. The main elements of Exploratory Data are finding out what happened, finding and finding new insights, and generating ideas and hypotheses from datasets that are analyzed or in the form of training datasets (Garousi et al., 2022). Figure 1 shows the process carried out in the exploration of data in this paper. The dataset is collected from the Youtube digital platform, and then Exploratory Data Analysis is carried out.

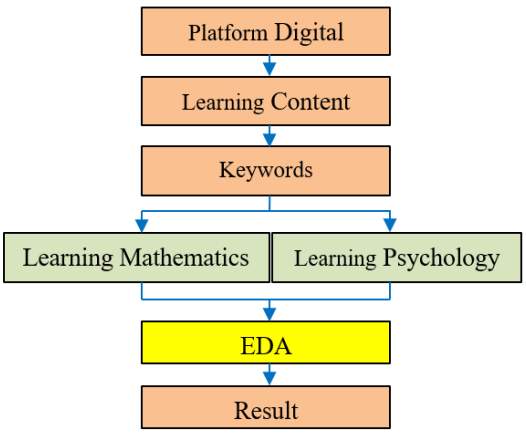

Figure 1. Data Exploration Process

Table 1. Duration and Views of Learning Content

Duration BM	Views BM	Duration BP	Views BP
9,25	2,243,668	26,27	1,017,999
9,03	7,353,202	11,13	3,903,910
4,48	533,958	9,46	1,389,913
5,40	1,260	9,34	766,534
4,29	2,244,007	5,06	32,253
5,13	449,157	11,17	204,791
5,28	1,751,895	9,12	115,362
4,26	322,986	9,01	1,353,664

Table 2. Duration and Views of Learning Content

No	Column	Non-null Count	Dtype
1	Duration BM	30 non-null	float64
2	Views BM	30 non-null	object
3	Duration BP	30 non-null	float64
4	Views BP	30 non-null	int64

3 Result and Discussion

After the dataset is processed and the Exploratory Data Analysis process is carried out, the results of the exploratory process are visualized with several forms of graphs.

The variables of the duration of the content of learning mathematics and psychology are processed by summing all the datasets in each column of the duration of the content of learning in the fields of mathematics and psychology. The total duration of the mathematics learning content amounted to 309 minutes, and the duration of the psychology learning content amounted to 404 minutes. Figure 2 shows the total duration of BM and BP.

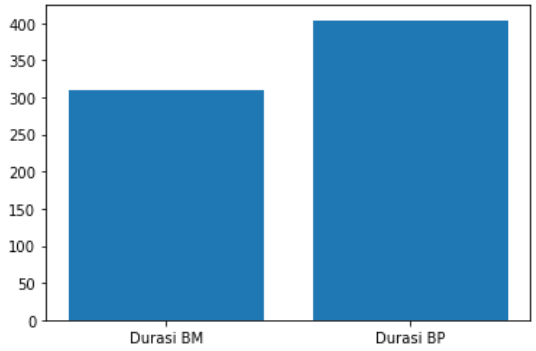

Figure 2. Total Duration of BM and BP

Figure 3 shows the number of views of mathematics learning content, as many as 32,770,853 views, and views of psychology learning content, as many as 19,939,180 views. The two pictures above show that out of 30 samples contained in duration and views. The duration variable shows that the duration of mathematics learning content has a smaller duration than psychology learning content. Meanwhile, in the views of the two learning content variables, views of mathematics learning content have a more significant number of views than

the number of views of psychology learning content.

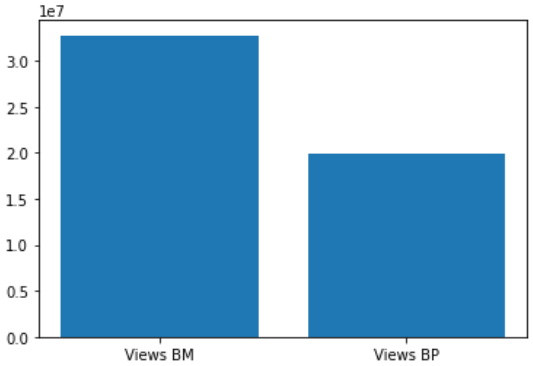

Figure 3. The sum of Views BM and Views BP

Table 3 shows all the data on the number, average, maximum and minimum in each column of the learning content dataset. The table shows variables in the number of durations of mathematics learning content as much as 309 minutes, an average duration of 10 minutes, a maximum duration of 38 minutes, and a minimum duration of 3 minutes. The total duration of psychology learning content is 404 minutes, the average duration is 13 minutes, the maximum duration is 32 minutes, and the minimum duration is 3 minutes. The variety of views in the two learning content consists of mathematics, consisting of 32,770,853 views, an average view of 1,170,388, a maximum of 7353202 views, and a minimum of 5 views. In psychology learning content, the number of views is 19,939,180, the average views are 664,639, the total views are 3,903,910, and the minimum views are 2,122.

In Figure 4, the results of the n_sampel processed in python programming that we show only in the number of durations of learning mathematics (BM) and the number of durations of learning psychology (BP) as well as n_sample average views of learning mathematics (BM) and views of learning psychology (BP).

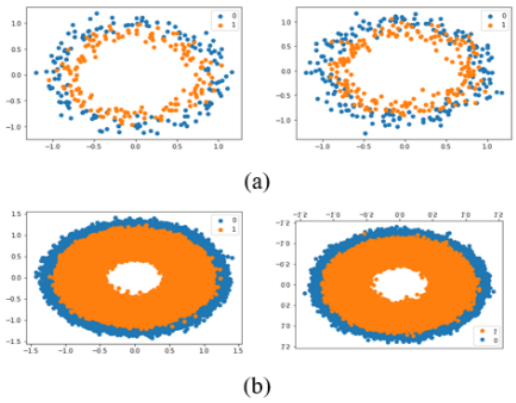

Figure 4. (a) Circles n_Sample sum duration BM vs. duration BP (b) Circles n_Sample average views BM vs. views BP

After doing some Exploratory Data Analysis process, we found new information related to learning content on the digital youtube platform. The duration variable shows the result that the total duration of the overall duration in mathematics learning content that includes the exact field is less than the psychology learning content included in the non-exact field. Datasets on exact planes can provide solutions to various nonlinear differential equations (Chen et al., 2022). Meanwhile, the overall number of views on mathematics learning content is more than the number of views on psychology learning content. The average duration of mathematics learning content is 309 minutes, less than the psychology learning content, which is 404 minutes. The average views of mathematics learning content are at 10 minutes, more than psychology learning content with 13 minutes. From the maximum value of mathematics learning content, it is more than psychology learning content, which is 38 minutes versus 32 minutes. The minimum value of the duration of

Table 3. Sum, Average, Max, Min

Variable	Sum	Average	Max	Min
Duration BM	309	10	38	3
Duration BP	404	13	32	3
Views BM	32770853	1170388	7353202	5
Views BP	19939180	664639	3903910	2122

the mathematics learning content and the minimum value of the psychology learning content is equal to the duration of 3 minutes. The maximum score for mathematics learning content is 7,353,202 views and 3,903,910 for the total views of psychology learning content.

Meanwhile, the minimum views of mathematics learning content are at 5 minutes, and 2,122 for the minimum views of psychology learning content. From the Exploratory Data Analysis that has been carried out, we also find different variations of the analysis's variables. This is also found in studies such as Whitelock-Wainwright et al. (2021), which analyzes the evaluation of student expectations of learning services. The findings from their study show that students' expectations regarding the ethical and privacy variables of learning services are consistent across all groups. However, the expectations of the service's features vary considerably.

4 Conclusion

The variables carried out exploratory data analysis both from duration and views, and we found differences. From these two variables, the content of mathematics learning included in the field of exact sciences has a higher weight on the number of views. Meanwhile, the content of psychology learning included in the non-exact field has a higher weight on the number of durations. In the future, research can be done by comparing and looking for correlation values between the two variables and adding variables that can be analyzed, such as the number of likes, subscribers, year of content creation, and making sentiment analysis of comments on learning content contained on the digital youtube platform. We can already explain the process and why we made this paper with several stages, starting from the introduction, materials, and methods used. The results and discussions we have comprehensively made. With the Exploratory Data Analysis, we can dig up data about learning content on digital platforms such as Youtube. From the collecting dataset that we have made, showing a vast number of views is undoubtedly the key to equitable distribution of information and knowledge for all users. More innovation and creating learning content are expected to encourage increased human development.

Authors Contributions

Mambang, planning and evaluating the entire manuscript, Haniffah Sri Rinjani collected the dataset and presented the dataset in the form of a table. Muhammad Zulfadhilah, Finki Dona Marleny, Septian Eka Prastya and

Subhan Panji Cipta, made data analysis using python programming language. All authors are thoroughly involved in reviewing the manuscript process.

Acknowledgement

The author gratefully the support from the Information Technology Department at Sari Mulia University and the Informatics Department at the Muhammadiyah University of Banjarmasin.

Referensi

Adeniyi, M. O., Ekum, M. I., C, I., S, O. A., A, A. J., Oke, S. I. and B, M. M. (2020), 'Dynamic model of COVID-19 disease with exploratory data analysis', *Scientific African* **9**, e00477.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S2468227620302155*

Ahmadi, M., Mohd Osman, M. H. and Aghdam, M. M. (2020), 'Integrated exploratory factor analysis and Data Envelopment Analysis to evaluate balanced ambidexterity fostering innovation in manufacturing SMEs', *Asia Pacific Management Review* **25**(3), 142–155.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1029313217305699*

Baran, C. and Yilmaz Baran, S. (2021), 'Youtube videos as an information source about urinary incontinence', *Journal of Gynecology Obstetrics and Human Reproduction* **50**(10), 102197.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S2468784721001343*

Brandstetter, D., Yang, X., Lüftner, D., Tautz, F. S. and Puschnig, P. (2021), 'kMap.py: A Python program for simulation and data analysis in photoemission tomography', *Computer Physics Communications* **263**, 107905.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0010465521000461*

Chacon-Hurtado, J. and Scholten, L. (2021), 'Decisi-o-rama: An open-source Python library for multi-attribute value/utility decision analysis', *Environmental Modelling & Software* **135**, 104890.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1364815220309476*

Chen, H., Zhu, Q. and Qi, J. (2022), 'Further results about the exact solutions of conformable space–time fractional Boussinesq equation (FBE) and breaking soliton (Calogero) equation', *Results in Physics* **37**, 105428.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S2211379722001863*

Garousi, V., Cutting, D. and Felderer, M. (2022), 'Mining user reviews of COVID contact-tracing apps: An exploratory analysis of nine European apps', *Journal of Systems and Software* **184**, 111136.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0164121221002338*

Hammouchi, H., Cherqi, O., Mezzour, G., Ghogho, M. and Koutbi, M. E. (2019), 'Digging Deeper into Data Breaches: An Exploratory Data Analysis of Hacking Breaches Over Time', *Procedia Computer Science* **151**, 1004–1009.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1877050919306064*

Ide, J. S., Li, H.-T., Chen, Y., Le, T. M., Li, C. S., Zhornitsky, S. and Li, C.-S. R. (2020), 'Gray matter volumetric correlates of behavioral activation and inhibition system traits in children: An exploratory voxel-based morphometry study of the ABCD project data', *NeuroImage* **220**, 117085.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1053811920305711*

Li, C., Zhang, S., Garza, T. and Jiang, Y. (2021), 'Data of the constructivist practices in the learning environment survey from engineering undergraduates: An exploratory factor analysis', *Data in Brief* **39**, 107522.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S2352340921007988*

Meyer, D. W. (2021), 'Netflow Python library – A free software tool for the generation and analysis of pore or flow networks', *MethodsX* **8**, 101592.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S2215016121003824*

Sorg, K. and Khobzi, H. (2022), 'A decade of the Swiss electronic vaccination Record: Some insights based on an exploratory data analysis', *International Journal of Medical Informatics* **158**, 104660.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1386505621002860*

Tanskanen, S.-K. (2021), 'Fragmented but coherent: Lexical cohesion on a YouTube channel', *Discourse, Context & Media* **44**, 100548.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S2211695821000842*

Whitelock-Wainwright, A., Tsai, Y.-S., Drachsler, H., Scheffel, M. and Gašević, D. (2021), 'An exploratory latent class analysis of student expectations towards learning analytics services', *The Internet and Higher Education* **51**, 100818.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1096751621000270*

Zanovello, U., Seifert, F., Bottauscio, O., Winter, L., Zilberti, L. and Ittermann, B. (2022), 'CoSimPy: An open-source python library for MRI radiofrequency Coil EM/Circuit Cosimulation', *Computer Methods and Programs in Biomedicine* **216**, 106684.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S0169260722000694*