

Application of Data Mining to Predict Student Satisfaction in Academic and Non-Academic Services at the Universitas Terbuka through Social Media-YouTube

Rhini Fatmasari¹ and Tora Akadira²

{riens@ecampus.ut.ac.id¹, toraakadira@ecampus.ut.ac.id²}

Universitas Terbuka, Indonesia^{1,2}

Abstract. Universitas Terbuka (UT) is a university with social media platforms such as Facebook, Twitter, Instagram, Google Plus, and LinkedIn to introduce and disseminate information about UT. This platform is also used as an education for students and the public. Posts on social media are fascinating to observe. The post contains various comments. The data are grouped to obtain specific patterns. It analyzes posts on social media, especially YouTube, related to student satisfaction with academic and non-academic services at Universitas Terbuka. The analysis was carried out using Text mining with the Cross Industry Standard Process for Data Mining (CRISP-DM) method. The results of the study show that of the 7,705 posts on YouTube, the most common problems raised are related to tutorials (13.49%), exams (3.38%), teachers (5.25%), modules (3.71%), applications (3.41%); non-academic (2.87%); and scholarships (0.88%). Meanwhile, in each category, student posts were found to be related to (a) satisfaction and dissatisfaction with academic and non-academic services provided by UT, (b) explaining the procedure for studying at UT, (c) announcements about the UT program, and (d) greetings to other students.

Keywords: social media, academic and non-academic services, data meaning

1. Introduction

Social media is a tool to socialize with each other online. Social media supports interaction using web-based technologies that turn communication into interactive dialogue. The largest social media networks include Facebook and Twitter [1]. Traditional media uses print and broadcast media, but social media uses internet networks. Social media invites users to participate by providing open contributions and feedback, commenting, and sharing information in a fast and unlimited time. Social media use, which is relatively easy to apply and reaches a very large mass, makes social media very popular. Various posts can be seen on the user's wall. Posts that attract attention are expressions of satisfaction or dissatisfaction of an individual with the events they experience. These posts often provoke responses from other

users, both positively and negatively. Universitas Terbuka has provided platforms on social media such as Facebook and Instagram. Since 2013, UT has used Facebook, Twitter, Instagram, google plus, and LinkedIn to introduce and disseminate UT information and educate students and the public. Since it was first used, the development of followers for social media used by UT until 2019 has continued to increase.

The number of followers on the UT social network at <http://www.facebook.com/univterbuka>, [@UnivTerbuka](http://twitter.com/univterbuka), <http://google.com/univterbuka>, and <https://www.linkedin.com/school/universitas-open> which can be seen in Figure 1.



Figure 1. Number of UT Social Media followers in 2018 and 2019
(Source: Chancellor's Report 2019)

Student activities through social media have increased because currently, UT students are younger. 43.89% of UT students are in the age range below 25 years. The statistical report noted that Indonesia's most social media users in 2020 were aged 25-34 years. Male users were 20.6%, and female users were 14.8% (<https://databoks.katadata.co.id/datapublish/2020/>).

Table 1. Number of UT Students by Age

| Age Range | Amount | Percentage |
|-----------|---------|------------|
| < 25 | 136.512 | 43,89% |
| 25-29 | 62.799 | 20,19% |
| 30-34 | 40.887 | 13,15% |
| 35-39 | 33.728 | 10,84% |
| 40-44 | 19.443 | 6,25% |
| > 44 | 17.659 | 5,68% |
| Total | 311.028 | 100% |

Posts and data, and comments on social media are very interesting. These data can be grouped to obtain certain patterns. Data mining is a data processing method to find hidden patterns in the data. The results of data processing with this data mining method can be used to make decisions in the future. Data mining is also known as pattern recognition [2]. Data mining is a large-scale data processing method. Therefore data mining has an important role in industry, finance, weather, science, and technology. Generally, data mining research discusses methods such as clustering, classification, regression, variable selection, and market basket analysis [3]. Data mining is discovering useful new correlations, patterns, and trends by mining many data repositories. Data processing uses pattern recognition technology such as statistics and mathematical techniques. Data mining is also known as knowledge discovery in

database (KDD) or pattern recognition [4]. Clustering divides data into classes or clusters based on similarity [5]. It is related to the variety of data that can be obtained through social media and the ability of data mining to analyze these data patterns.

Student posts on social media are very interesting to observe. These comments describe students' attention to academic and non-academic services provided by the Universitas Terbuka. These comments can also describe the quality of service and student satisfaction. Student comments on social media as Big Data in the Data Mining research model are very valuable data. The data can be analyzed and used as a tool for making decisions. The researchers are interested in researching student satisfaction with UT academic services, both academic and non-academic services, and problems related to those services. Based on this phenomenon, it is very important to do this research so that this valuable data source can be used effectively .

2. Methods

This research was conducted by analyzing posts on UT Social Media, then interviewing several Distance Learning Program Unit UT samples for clarification. The sample selection of the Distance Learning Program Unit (DLPU) UT used a purposive sampling method. The sampling selection of DLPU UT considered the representation of DLPU UT Large, DLPU UT Small, and DLPU UT Western, Central, and Eastern Regions.

2.1 Research design

This research method refers to research conducted by Achmad Bayhaqy [6]. The research methodology uses the Cross-Industry Standard Process for Data Mining (CRISP-DM) method, which consists of six stages.

- a. Business Understanding. Answering the need to solve a problem using Data Mining, in this case, is the satisfaction of UT students with UT's academic and non-academic services written by students on social media.
- b. Data Understanding. Initial data collection is done from social media, namely Facebook, Twitter, and Instagram UT. The data used are posts and tweets from UT students for five months using the "Data Scraper" application. "Data Scraper" is a plug-in in the Chrome browser that can collect each user's tweets and posts. The data structure is taken from Facebook, Twitter, and Instagram UT consists of several columns, namely Username, Fullname, Date, First Hashtags, Alt Hashtags, Tweet (Message Text), Retweet, Retweet (w/ Text), Likes, Likes (w/ Text).), URL Paths. The column used is only "Tweet (Message Text)."
- c. Data Preparation Models. The next stage is to model the calculation model with the scenario stages that will be processed, then the 10-Fold Cross Validation test.
- d. Evaluation. The evaluation stage evaluates the performance of the modeling and calculations carried out by measuring the performance of the classification algorithm used. In this study, performance is measured using accuracy, precision, recall, F1-Score, and Area Under Curve (AUC), which will be displayed in the form of a ROC curve and word cloud to visualize the spread of positive and negative words that usually appear from TF-IDF. These results will show the use of the most appropriate algorithm for the model that has been proposed previously.
- e. Deployments. The application of the model into the application with the best evaluation value obtained at the Evaluation stage. The application is to be developed using the python programming language and the required libraries [7]

3. Result and Discussion

Student posting data on YouTube is taken within five months (March to August 2022). The data analyzed amounted to 7,705 posts. Student posts are grouped into 8 (eight) categories of comments: (1) teaching material problems; (2) tutorial problems; (3) scholarships; (4) lecturers and tutors; (5) the implementation of the exam, both the TAP exam and the final exam; (6) e-learning applications; (7) non-academic services at UT Center and UPBJJ UT; and (8) other comments.

The stages of data processing are presented in Figure 2

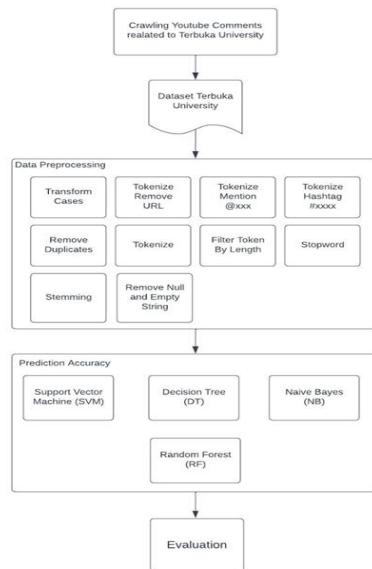


Fig. 2. The stages of data processing

The study results show that out of 7,705 posts on Youtube, the categories of problems posted are shown in Table and Figure 2 below.

Table 2. Categories of Posts on YouTube

| Category Posts | Amount | Percentage |
|----------------|--------------|------------|
| Module | 193 | 2.73 |
| Tutorial | 701 | 9.91 |
| Scholarship | 46 | 0.65 |
| Teacher/ Tutor | 273 | 3.86 |
| Exam | 338 | 4.78 |
| App | 177 | 2.50 |
| Non-Academic | 149 | 2.11 |
| Other Comments | 5.198 | 73.47 |
| Total | 7.075 | 100 |

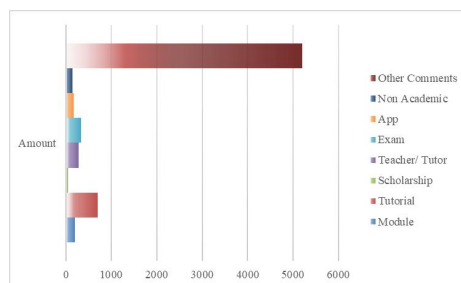


Fig. 3. Categories of Posts on YouTube

The data in Table 2 shows that most posts are related to problems with other UT academic and non-academic services. The problems often discussed on the Youtube wall are more related to the experiences of seniors or their friends who have previously studied at UT. For example

- the desire to study online
- get acquainted with friends who have just studied at UT
- Share experiences while studying at UT
- greet friends who are in the same department
- interested in studying at UT after receiving an explanation from a friend
- how to make payment
- encouraging his friends who will enter UT
- ask about accreditation
- ask for the address of DLPU UT in the area
- Interested in studying at UT after watching celebrity youtube content; for example, someone wrote:
"At first, I was interested in studying at UT, where I could study online after watching Raditya Dika's podcast".

In each category of student posts related to (a) satisfaction and dissatisfaction with academic and non-academic services provided by UT, (b) explaining the procedures for studying at UT, (c) announcements about the UT program, and (d) greetings to other students. Student comments regarding online and face-to-face tutorials are presented in Table 3.

Table 3. Categories of Posts on YouTube about Tutorials

| Category Posts | Amount | Percentage |
|------------------------------|------------|-------------|
| How to activate Tutor | 121 | 24.59% |
| Explain to friends | 265 | 53.86% |
| Notifications to friends | 9 | 1.83% |
| Say hello to students/Tutors | 1 | 0.20% |
| Dissatisfied Student | 57 | 11.59% |
| Satisfied Student | 39 | 7.93% |
| Total | 492 | 100% |

Table 3 shows that after watching the material about UT on YouTube, there were many questions about the material being broadcast. Other students answered these questions. 53.86% of students who also watched the same material gave explanations to their friends. The comments are also about how to activate Tutor (25.59%).

Table 3 also shows that 7.93% of students expressed satisfaction with the Online Tutorial and Face-to-face Tutorial services because they can increase their score for the Final Semester Exam. Meanwhile, 11.59% of students expressed dissatisfaction with the Online Tutorial and Face-to-Face Tutorial services.

4. Conclusions

Social media, such as YouTube, is one of the media used by Universitas Terbuka (UT) to explain UT programs to students. YouTube is also used to provide socialization and promotion about UT. Based on the results of research on Predict Student Satisfaction in Academic and Non-Academic Services at the Universitas Terbuka through Social Media YouTube, it can be concluded that after watching YouTube shows, 73.47 students made comments related to their experiences with UT. The comments included wanting to study online, being interested in studying at UT after receiving an explanation from friends, encouraging friends who will enter UT, accreditation, and DLPU UT address in the area. Some students are interested in studying at UT after watching youtube content from a celebrity.

These comments are very interesting to observe so that UT can get an idea of the public's interest in studying at UT and student satisfaction after studying at UT. The most comments on youtube (73.47%) are not academic problems, but comments to share experiences between fellow UT students.

5. References

- [1] Wisti Aristika, Wira Jaya Hartono.: Penerapan Clustering K-Means Untuk Menentukan Pengaruh Media Sosial Facebook Terhadap Usaha Mikro, Kecil Dan Menengah (UMKM) Di Kecamatan Pekanbaru Kota. *Jurnal Ilmu Komputer dan Bisnis*, Volume 11, Nomor 1, Mei (2020)
- [2] A.P. Windarto.: Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method. *Int. J. Artif. Intell. Res.*, Vol. 1, no. 2. pp. 26–33 (2017)
- [3] J. O. Ong.: Implementasi Algoritma K-means clustering untuk menentukan strategi marketing president university. *J. Ilm. Tek. Ind.*, vol. vol.12, no, no. juni, pp. 10–20 (2013)
- [4] Fatmawati, Kiki & Windarto, Agus.: Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi. *Computer Engineering, Science and System Journal*. 3. 173. 10.24114/cess.v3i2.9661 (2018)
- [5] Khotimah, T.: Pengelompokan Surat Dalam Al Quran menggunakan Algoritma K-Means. *Jurnal Simetris*, 5(1), pp. 83-88 (2014)
- [6] Achmad Bayhaqy.: Analisa Sentimen dengan Pelabelan Otomatis menggunakan VADER tentang Islamophobia pada Media Sosial. Thesis, August 2019 DOI: 10.13140/RG.2.2.19304.08966. STMIK Nusa Mandiri Jakarta (2019) <https://www.researchgate.net/publication/335380972>
- [7] Han, J., & Kamber, M.: *Data Mining Concept and Tehniques*. Morgan Kauffman, San Fransisco (2006)