JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER

p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 451-458
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.403

Analysing User Reviews with ETL using Pentaho Data Integration

Tri Luhur Indayanti Sugata[1,*], Rafika Rahmawati[1], Tri Puspa Rinjeni[1], Virdha Rahma Aulia[1], Prasasti Karunia Farista Ananto[1], Iqbal Ramadhani Mukhlis[1]

[1]Faculty of Computter Science, Information System, Universitas Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia
Email: [1]tri.luhur.fasilkom@upnjatim.ac.id, [2]rafika.rahmawati.fasilkom@upnjatim.ac.id, [3]virdha.rahma.fasilkom@upnjatim.ac.id, [4]iqbal.ramadhani.fasilkom@upnjatim.ac.id, [5]prasasti.karunia.fasilkom@upnjatim.ac.id, [6]puspa.rinjeni.fasilkom@upnjatim.ac.id
(*Email Corresponding Author: [1]tri.luhur.fasilkom@upnjatim.ac.id)
Received: June 23, 2025. | Revision: July 7, 2025 | Accepted: July 7, 2025

Abstract

User reviews are a crucial element in guiding the continuous improvement of mobile applications for developers. This research aims to utilize Extract, Transform, Load (ETL) techniques using Pentaho Data Integration to analyze user reviews of government mobile applications which is 'Sentuh Tanahku', focusing on improving service quality through actionable data insights. The ETL process involves collecting and cleaning data from the Google Play Store to derive valuable insights that inform recommendations for app improvement. After data extraction, text preprocessing steps, such as cleansing, case folding, and keyword filtering, were applied to prepare the data for analysis. By categorizing user reviews into key aspects such as user interface, performance enhancement, bug fixes, security, compatibility and feature development this research enables the identification of most frequently discussed and complained about by users. The output of this research includes a structured dataset in Excel format. By demonstrating the effectiveness of ETL and text analysis in transforming unstructured user reviews into strategic insights, this research contributes to utilizing Pentaho Data Integration as an alternative and effective tool for processing and analyzing user reviews.

Keywords: Pentaho Data Integration, User Review, ETL, Sentuh Tanahku, Data

1. INTRODUCTION

Mobile applications significantly improve public access to governmental services, such as the "Sentuh Tanahku" application, which is accessible via Android and iOS platforms and can be accessed from anywhere [1]. The availability of the Sentuh Tanahku application accelerates services and makes it easier to obtain land service information [2]. User reviews contain a wealth of information that influences app performance and have been proven to significantly increase app revenue [3].

However, manually evaluating user reviews can be labor-intensive and inefficient due to the large amount of unstructured data found in app marketplaces such as the Google Play Store. To address this, various techniques have been proposed for managing and analyzing user reviews including sentiment analysis, topic modeling or using deep learning techniques. Sentiment analysis is an automated process for extracting, analyzing, and processing textual data to get information included in an opinion sentence [4]. Sentiment analysis such as Random Forest [5] and Support Vector Machine (SVM) [6] help to classify user reviews from mobile applications into positive, neutral and negative sentiment.

Topics modelling methods like Latent Dirichlet Allocation (LDA) is a probabilistic generative model capable of uncovering underlying semantic themes within review documents[7]. More recently, deep learning techniques are also used such as CNN and Bi-LSTM, which can automatically extract classification features and understand semantic meaning and context [8]. While these approaches being effective, they are often require complex and computationally demanding.

To address this gap, this research propose the use of Extract, Transform, Load (ETL) techniques with Pentaho Data Integration (PDI) as simpler and low-cost an alternative. ETL offers an efficient approach to processing large data. One widely used ETL tool is Pentaho Data Integration, which supports the processing and analysis of data from multiple sources [9]. Prior studies have successfully implemented ETL using Pentaho across various domains. In commercial domain, Pentaho had been used to manage a sales data to accelerate sales performance [10] and find product categories with the greatest sales [11]. In application feedback context, also be utilized to extract, cleanse and load user comments into the data warehouse [12]. In public sector and governmental settings, it has been used for data integration processes [13] and analytical reports [14]. Various studies have shown that Pentaho is a reliable ETL platform capable of processing unstructured data into meaningful information. Therefore, its capabilities make it a suitable choice to support this research.

This research aims to systemati analyze user reviews of the "Sentuh Tanahku" application by categorizing reviews into distinct aspects: user interface, performance enchancement, bug fixes, security, compatibility, and feature development. The contribution of this research is the demonstration ETL's effectiveness in converting unstructured text data into actionable insights, offering a lightweight and cost-effective solution especially suitable for government institutions with limited data science resources.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 451-458
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.403

2. RESEARCH METHODOLOGY

The research methodology employed in this research is illustrated in Figure 1, showing the ETL process. Figure 1 provides a graphical representation of the ETL workflow from initial data scraping (extract), through preprocessing stages (transform) using Pentaho Data Integration, and finally loading processed data into Excel for analysis.

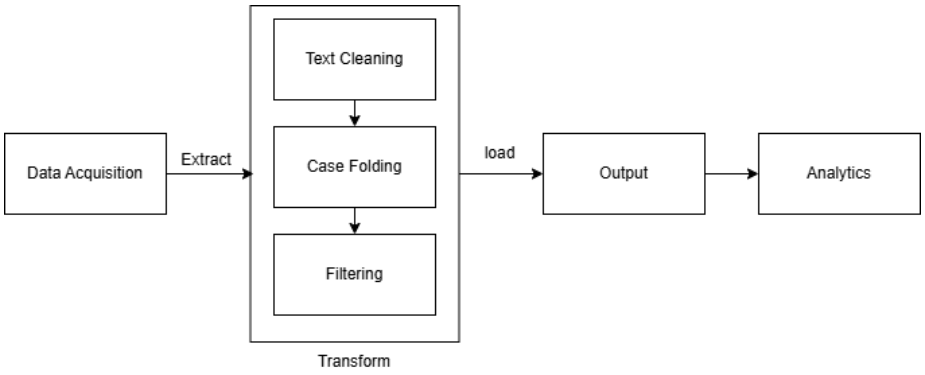

Figure 1. Research Method

2.1 Extraction

In ETL Technique, the first step is extraction. The extraction phase in this step begins by using scraping technique to obtain user reviews of Sentuh Tanahku application from Google Play Store. This is done by using python programming and *google_play_scraper* library. A total of 14.143 user reviews were successfully collected. These reviews include valuable metadata such as review content, rating score, app version, reviewer username, and time.

2.2 Transformation

Online texts often contain noise, such as HTML tags, scripts, advertisements, and unimportant words that have minimal impact on their overall meaning [15]. During transformation, data will be processed using text cleaning (removing unwanted symbols, emojis, hashtags, mentions, and any other irrelevant characters, leaving only alphanumeric characters [16]), case folding (normalizing text to lowercase and differences between uppercase and lowercase are treated as the same [17]) and keyword filtering based based on predefined categories (e.g., bug fixes, performance enhancement). String matching is a basic technique used for categorizing user reviews based on predefined aspects such as bug reports, user experiences, feature requests, or ratings [18]. A stop words removal step was also implemented using the Sastrawi library imported via Pentaho's Text File Input step. These preprocessing steps help normalize the data [19] and prepare text for accurate analysis.

2.3. Loading

In the Loading phase of ETL, once the data has been filtered by specific aspects, the processed information is stored in an Excel file to facilitate further analysis and reporting. This final phase involves transferring the transformed data into the Excel format, where stakeholders or developers can easily access and interpret the information. After filtering data by aspects like perbaikan bug, peningkatan performa, keamanan, kompatibilitas, antarmuka pengguna and pengembangan fitur, each category data is loaded into the Excel file. The Excel file will contain 2 sheets with different information. This structure enables data analysts, developers or even owners to examine each aspect independently, perform additional analyses, visualize trends such as rating, and prioritize areas needing attention.To get better insights, an additional step involves adding stop words to the filtering process by eliminating common but non-informative terms. Subsequently, a word count analysis is performed to identify frequently occurring words in user reviews. By loading the data into Excel, teams can quickly generate visual reports and charts, making it easy to monitor, and help to identify which problems need to be solved based on user review for app improvement.

3. RESULTS AND DISCUSSION

This section presents the overall ETL (Extract, Transform, Load) process designed using Pentaho Data Integration (PDI). The workflow consists of extracting data from CSV file input, transformation steps and load them into an Excel Format. This Figure 2 below illustrates the complete workflow applied in this research.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 451-458
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.403

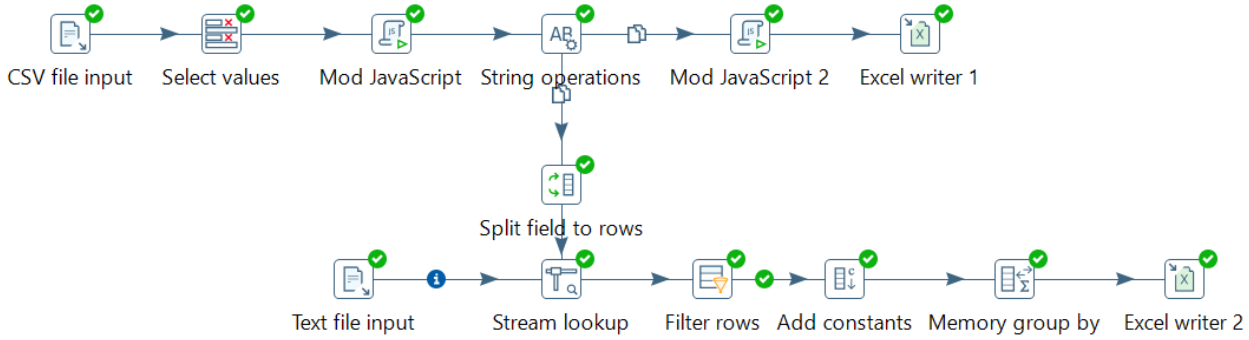

Figure 2. Pentaho workflow

3.1 Extraction Output

In Pentaho, CSV file input step is used to extract data obtained from the Google Play Store. The result of this phase is a raw dataset consisting of 14,143 records of user reviews from google play store. Data then extracted to specifically target necessary fields such as rating, username and user review content. This data extraction ensures that the key elements needed for analyzing user reviews are collected, so that it only focuses on specific aspects that users frequently mention in their reviews.

3.2 Text Cleaning

Following extraction, the text cleaning phase is important to prepare the data for meaningful analysis because raw user reviews often contain various characters. This is necessary for ensuring consistency in the data and avoiding any noise that could interfere in the filtering and analysis stages. The text cleaning process makes the data more general and readable by applying cleaning functions. This process is implemented using the *Modified JavaScript* step in Pentaho. Table 1 shows the results after the text cleaning process.

Table 1. Text Cleaning Result

Original Text	After Text Cleaning
Info lengkap, jelas & kereeen 👍	Info lengkap jelas kereeen
fitur pertanyaan belum bisa/belum aktif	fitur pertanyaan belum bisa belum aktif
Tolong Sempurnakan lagi ! Terimakasih	Tolong Sempurnakan lagi Terimakasih

3.3 Case Folding

Once we have removed unwanted symbols and standardized the text content, case folding is applied to further normalize the data by converting all text to lowercase. In Pentaho, this case folding process is performed using the String Operations step. This step is necessary to ensure that keywords are not missed due to capitalization differences, as it eliminates distinctions between uppercase and lowercase letters as shown in Table 2 below.

Table 2. Case Folding Result

Original Text	After Case Folding
Apk Sentuhku Sangat Membantu	apk sentuhku sangat membantu
SEMOGA SHM CEPAT JADI, AMIIN	semoga shm cepat jadi amiin
Login pakai akun gak bisa setelah pakai Ktp baru bisa.	login pakai akun gak bisa setelah pakai ktp baru bisa

3.4 Keyword Filtering

After data preprocessing, filtering based on specific aspects using string matching allows us to categorize user reviews into focused areas that highlight different dimensions of the app's quality and functionality. In this research, we define five main aspects in addition to organize user reviews, each representing a distinct area of concern or improvement for the mobile app: *perbaikan bug* (bug fixes), *peningkatan performa* (performance enhancement), *keamanan* (security), *kompatibilitas* (compatibility), *antarmuka pengguna* (user interface) and *pengembangan fitur* (feature development). By using these predefined keywords, we can simply sort through user reviews, pinpointing specific issues within each category as shown in Table 3.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 451-458
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.403

Table 3. Predefined Keywords

Aspect	Keywords
Perbaikan Bug	bug, error, masalah, gagal, crash, tidak berfungsi, kerusakan, force close
Peningkatan Performa	performa, kecepatan, loading, respon, lambat, optimalisasi, perform
Keamanan	keamanan, privasi, proteksi, data aman, security, sekuriti
Kompatibilitas	kompatibilitas, dukungan, versi, update, perangkat, sistem operasi, kompatibel, kompatibilitas
Antarmuka Pengguna	antarmuka, tampil, tampilan, UI, UX, desain, user interface, navigasi
Pengembangan Fitur	fitur, pengembangan, tambahan, baru, update, request, penambahan

The result of keyword filtering based on specific aspect shown in Figure 3. The bar chart presents the categorization of 14,143 user reviews into six predefined aspects based on keyword filtering. The most frequently mentioned category is *Perbaikan Bug*, which accounts for 916 reviews, followed by *Kompatibilitas* issues with 667 reviews, and *Pengembangan Fitur* with 387 reviews. *Peningkatan Performa* received 223 mentions, while *Antarmuka Pengguna* and *Keamanan* are the least discussed categories, with only 44 and 21 reviews respectively.

The remaining 11,885 reviews fall into the Others category, which includes feedback that did not match the defined keywords for the main six aspects. The analysis reveals that the majority of user concerns focus on technical aspects, particularly bug fixes and compatibility issues. This indicates that Sentuh Tanahku application issues need to be solved. Furthermore, a significant portion of reviews falls outside the predefined categories, indicating the presence of diverse user needs. This highlights the limitation of the current keyword-based categorization approach, which may overlook contextual. Therefore, more in-depth analysis or advanced techniques are needed to support future updates and ensure more targeted and comprehensive improvements

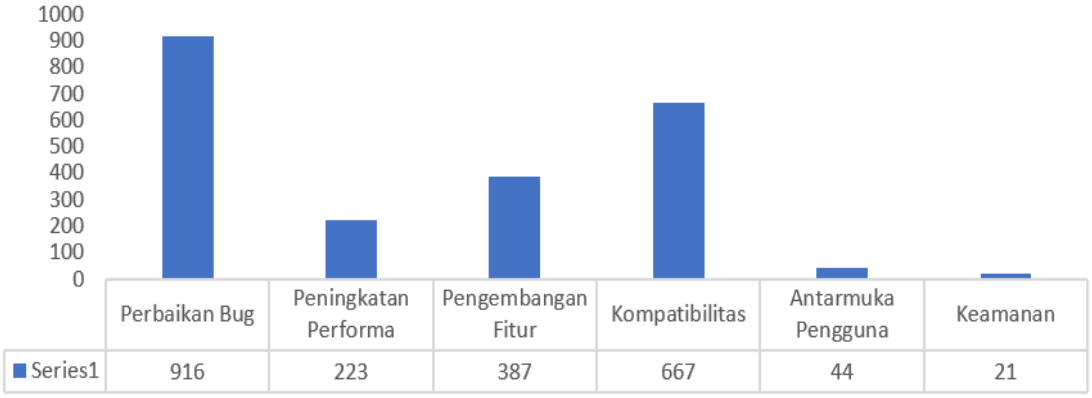

Figure 3. Keyword Filtering Result

3.5 Stopword Removal

The transformation results often still contain many irrelevant words, such as prepositions and others, which can overshadow important words. Therefore, these need to be removed, typically using a predefined stopwords list [20]. The stopword list is obtained from the Sastrawi library. In Pentaho, this list was imported using the *Text File Input* step, which is then compared using the *Stream Lookup* step. The output from the *Stream Lookup* step can be seen in Figure 4. Entries in the stopwords column with values other than null will be removed from the list.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 451-458
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.403

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

First rows | Last rows | Off

#	userName	content	score	haa	stopwords
1	Ismani Iis	Baik	5	baik	baik
2	M.TOHAN RIYANTO	Sangat...	5	sangat	sangat
3	M.TOHAN RIYANTO	Sangat...	5	membantu	<null>
4	Anca Herdiansyah	Tidak ...	1	tidak	tidak
5	Anca Herdiansyah	Tidak ...	1	menarik	<null>
6	Anca Herdiansyah	Tidak ...	1	lagi	lagi
7	Anca Herdiansyah	Tidak ...	1	dikarnakan	<null>
8	Anca Herdiansyah	Tidak ...	1	swaploting	<null>

Figure 4. Stream Lookup Step Result

3.6 Word Cloud

To gain deeper insight into the user reviews, word clouds can be used. A word cloud is an effective data visualization tool for quickly identifying the content or topics of text documents without the need to read them entirely [20] . To make a word cloud, we need to calculate the frequency of words that appear in user reviews. Then Add constant and Memory group by step is used in Pentaho.

	A	B	C	D	E
1	userName	content	score	hasil	kategori
2	Ismani Iis	Baik	5	baik	Lainnya
3	M.TOHAN RIYANTO	Sangat membantu	5	sangat membantu	Lainnya
4	Anca Herdiansyah	Tidak menarik lagi d	1	tidak menarik lagi dikarnakan swaplotir	Perbaikan Bug
5	Dede Hermansah	Bagus	5	bagus	Lainnya
6	Kartin Karim	Ok	5	ok	Lainnya
7	Cahaya Tahajud	Mantap	5	mantap	Lainnya
8	Ir. Toto Hartanto	Info lengkap, jelas &	5	info lengkap jelas kereeen	Lainnya
9	Jaka Suherman	Bisa membantu infc	3	bisa membantu informasi	Lainnya
10	PT. ANDRIA BUANA JAYA	Tidak ada plot, moh	1	tidak ada plot mohon untuk tampilkan p	Lainnya
11	Arief Rakhman	Sangat membantu	5	sangat membantu	Lainnya
12	burhan han	User frendly	5	user frendly	Lainnya
13	Jalalludin Ishak	Mantap	4	mantap	Lainnya
14	Ratna Watty	Tingkatkan profesio	4	tingkatkan profesionalitas kerja	Lainnya
15	Ahmad Saiful	Bgmn je aplikasi ini	1	bgmn je aplikasi ini tidak ada muncul ga	Lainnya
16	Vika Prahastuti S.H	kecewa bgt, minta t	3	kecewa bgt minta tlg ga di respon datar	Peningkatan Performa
17	Hnd channel	Udah punya akun tp	1	udah punya akun tp login selalu tidak b	Lainnya
18	New Normal	Kerenn	5	kerenn	Lainnya

Sheet1 | Sheet2

Figure 5. Sheet1 in Excel Output

The final output will be saved in a single Excel file containing two sheets. Sheet 1 is result from Excel Writer 1 step which stores filtered user review data contain username, user comment, rating and with their categorized aspects as shown in Figure 5, while Sheet 2 is result from Excel Writer 2 step which records the most frequently occurring words from the user reviews as shown in Figure 6.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 451-458
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.403

	A	B	C
1	word	frequency	
2	aplikasi	2988	
3	membant	2262	
4	tanah	1275	
5	bagus	1236	
6	gak	1081	
7	nya	957	
8	sertifikat	879	
9	daftar	820	
10	verifikasi	819	
11	login	774	
12	ga	756	
13	lokasi	721	
14	mantap	712	
15	email	708	
16	update	666	
17	cek	656	
18	masuk	651	

Sheet1 Sheet2

Figure 6. Sheet2 in Excel Output

After getting the word and its number of occurrences, a word cloud can also be created to get better insight. Figure 7 represents the word cloud generated from the Excel output in Sheet 2. Where the most frequently discussed words in the user reviews include "aplikasi", "membantu", "tanah", "bagus", “sertifikat” and others. The word cloud revealed prominent keywords like "sertifikat," suggesting a significant administrative focus in user concerns. This highlights the need for addressing administrative functionalities in future application updates.

Figure 7. Word Cloud

3.7 User Ratings

The analysis also included a breakdown of the rating scores. From the Excel output Sheet 1, a chart is generated to get better understanding about how much users rate the application as shown in Figure 7. From the chart, it can be concluded that many users gave a rating of 5, followed by a rating of 1. This indicates that users are satisfied with the Sentuh Tanahku application. However, the significant number of ratings of 1 also shows that some users are dissatisfied, indicating that improvements need to be made to address the issues.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 451-458
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.403

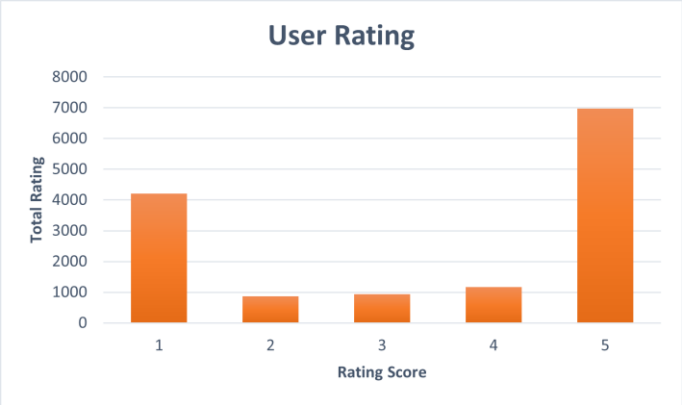

Figure 8. Total User Rating

4. CONCLUSION

This research successfully implemented Extract, Transform and Load (ETL) using Pentaho Data Integration to analyze user reviews of the Sentuh Tanahku government mobile application. The ETL process effectively extracted data from the Google Play Store, applied text preprocessing steps such as cleansing, case folding, and keyword filtering, categorized user reviews into key aspects and stopwords removal. From the result, user review mainly discusses about technical aspects, particularly bug fixes and compatibility issue. Which indicate that the developer needs to resolve the issues related to the Sentuh Tanahku application. The main contributions of this study include the application of ETL techniques to process user reviews using Pentaho. The novelty of this research lies in its lightweight, low-cost approach using open-source tools, which is suitable for government institutions with limited data science resources. This study is limited to keyword-based categorization and lacks deeper sentiment scoring. Future research could be employed by integrating sentiment analysis models to further deepen insights.

REFERENCES

[1] U. R. Sa'adah, Murwaniyah, D. I. Pradana, Masutiah, N. Panggabean, and Hamka, "APLIKASI SENTUH TANAHKU SEBAGAI INOVASI PELAYANAN PUBLIK DI KANTOR WILAYAH BADAN PERTANAHAN NASIONAL PROVINSI D.K.I. JAKARTA," *Jurnal Administrasi Bisnis Terapan*, vol. 5, no. 1, Dec. 2022, doi: 10.7454/jabt.v5i1.1037.

[2] Mariana Derlan Masia Harahap, F. Ferdinand, and Luluk Tri Harinie, "Pemanfaatan Aplikasi Sentuh Tanahku Guna Perbaikan Kinerja Layanan di Kantor Pertanahan Kota Palangka Raya," *Edunomics Journal*, vol. 4, no. 2, pp. 103–125, Jun. 2023, doi: 10.37304/ej.v4i2.10015.

[3] Z. Jiang, V. Liu, and M. Erne, "Examining the Usefulness of Customer Reviews for Mobile Applications," *Journal of Database Management*, vol. 35, no. 1, pp. 1–23, May 2024, doi: 10.4018/JDM.343543.

[4] H. Sujadi, "ANALISIS SENTIMEN PENGGUNA MEDIA SOSIAL TWITTER TERHADAP WABAH COVID-19 DENGAN METODE NAIVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE," *INFOTECH journal*, vol. 8, no. 1, pp. 22–27, Mar. 2022, doi: 10.31949/infotech.v8i1.1883.

[5] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," 2022. [Online]. Available: http://j-ptiik.ub.ac.id

[6] T. J. Firdaus, J. Indra, S. Arum, P. Lestari, and H. Hikmayanti, "SENTIMENT ANALYSIS OF THE SAMBARA APPLICATION USING THE SUPPORT VECTOR MACHINE ALGORITHM," vol. 5, no. 4, pp. 1183–1192, 2673, doi: 10.52436/1.jutif.2024.5.4.2673.

[7] G. Rosalinda, R. Santoso, and P. Kartikasari, "PEMODELAN TOPIK ULASAN APLIKASI NETFLIX PADA GOOGLE PLAY STORE MENGGUNAKAN LATENT DIRICHLET ALLOCATION," *Jurnal Gaussian*, vol. 11, no. 4, pp. 554–561, Feb. 2023, doi: 10.14710/j.gauss.11.4.554-561.

JURNAL KOMPUTER TEKNOLOGI INFORMASI SISTEM KOMPUTER
p-ISSN: 2962-3022 | e-ISSN: 2963-7104
Volume 4 No 2 September 2025
Page: 451-458
http://ejurnal.lkpkaryaprima.id/index.php/juktisi
DOI: https://doi.org/10.62712/juktisi.v4i2.403

[8] P. Bhuvaneshwari, A. N. Rao, Y. H. Robinson, and M. N. Thippeswamy, "Sentiment analysis for user reviews using Bi-LSTM self-attention based CNN model," *Multimed Tools Appl*, vol. 81, no. 9, pp. 12405–12419, Apr. 2022, doi: 10.1007/s11042-022-12410-4.

[9] A. D. Barahama and R. Wardani, "Utilization Extract, Transform, Load For Developing Data Warehouse In Education Using Pentaho Data Integration," *J Phys Conf Ser*, vol. 2111, no. 1, p. 012030, Nov. 2021, doi: 10.1088/1742-6596/2111/1/012030.

[10] I. Putu, W. Prasetia, I. Nyoman, and H. Kurniawan, "Implementasi ETL (Extract, Transform, Load) pada Data warehouse Penjualan Menggunakan Tools Pentaho," *TIERS Information Technology Journal*, vol. 2, no. 1, pp. 39–47, 2021, [Online]. Available: https://journal.undiknas.ac.id/index.php/tiers

[11] I. P. A. Eka Pratama and R. Bernard, "Analisa Kategori Barang dengan Penjualan Terbanyak dalam Jangka Waktu 3 Bulan Menggunakan Data Warehouse," *Jurnal ELTIKOM*, vol. 6, no. 1, pp. 65–78, Jan. 2022, doi: 10.31961/eltikom.v6i1.457.

[12] R. I. Alif, M. Idhom, and W. S. JS, "PENERAPAN TEKNIK ETL PADA KOMENTAR APLIKASI FLIP.ID DI APLIKASI PLAYSTORE DENGAN APLIKASI PENTAHO," *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, vol. 5, no. 2, pp. 1264–1272, Aug. 2024, doi: 10.46306/lb.v5i2.713.

[13] Purwita Sari, Lucky Indra Kesuma, Mira Afrina, and Dedy Kurniawan, "Pemodelan Integrasi Data Barang Milik Negara di Perguruan Tinggi Menggunakan Metode ETL (Extract, Transform, Load) dengan Pentaho," *The Indonesian Journal of Computer Science*, vol. 13, no. 5, Oct. 2024, doi: 10.33022/ijcs.v13i5.4424.

[14] N. W. S. Saraswati and N. M. L. Martarini, "Extract Transform Loading Data Absensi Stmik Stikom Indonesia Menggunakan Pentaho," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 19, no. 2, pp. 273–281, May 2020, doi: 10.30812/matrik.v19i2.564.

[15] E. Haddi, X. Liu, and Y. Shi, "The Role of Text Pre-processing in Sentiment Analysis," *Procedia Comput Sci*, vol. 17, pp. 26–32, 2013, doi: 10.1016/j.procs.2013.05.005.

[16] S. Malgaonkar, S. A. Licorish, and B. T. R. Savarimuthu, "Prioritizing user concerns in app reviews – A study of requests for new features, enhancements and bug fixes," *Inf Softw Technol*, vol. 144, Apr. 2022, doi: 10.1016/j.infsof.2021.106798.

[17] S. Pradha, M. N. Halgamuge, and N. Tran Quoc Vinh, "Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data," in *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, IEEE, Oct. 2019, pp. 1–8. doi: 10.1109/KSE.2019.8919368.

[18] W. Maalej, Z. Kurtanović, H. Nabil, and C. Stanik, "On the automatic classification of app reviews," *Requir Eng*, vol. 21, no. 3, pp. 311–331, Sep. 2016, doi: 10.1007/s00766-016-0251-9.

[19] K. Kmgs and R. Rahm, "Learning to Use Normalization Techniques for Preprocessing and Classification of Text Documents," 2022.

[20] Y. Kalmukov, "USING WORD CLOUDS FOR FAST IDENTIFICATION OF PAPERS' SUBJECT DOMAIN AND REVIEWERS' COMPETENCES 15." [Online]. Available: www.compsystech.org