e-ISSN: 3108-9828 | p-ISSN: 3109-3981
Volume 3 No 1 May 2026 | Page: 48–56
DOI: https://doi.org/10.64803/jocsaic.v3i1.182

Journal of Computer Science Artificial Intelligence and Communications

Journal Homepage: https://journals.raskhamedia.or.id/index.php/jocsaic

Web Scraping Design for Text Data Acquisition on Platform X in Adolescent Sexual Deviant Behavior Research

Yanti Yusman[1*], Noor Anida Zaria Mohd Noor[2]

[1]Faculty Sain Komputasi dan Kecerdasan Digital , Program Study Sistem Informasi ,Universitas Pembangunan Panca Budi, Medan, Indonesia
[2]Faculty of Computer Science and Meta Technology, Universiti Pendidikan Sultan Idris, Tanjong Malim-Perak, Malaysia
Email : [1*]yantiyusman@gmail.com

ARTICLE INFO

Article history:

Received May 19, 2026
Revised May 25, 2026
Accepted May 28, 2026
Publish May 31, 2026

ABSTRACT

The increasing use of social media among adolescents has generated vast amounts of digital textual data that can be utilized to investigate online behavioral phenomena, including discussions related to sexual behavior. As one of the most active social media platforms, X provides a rich source of publicly available textual information that reflects users' interactions, opinions, and behavioral expressions. However, obtaining high-quality textual data for behavioral research requires a systematic, transparent, and reproducible data acquisition process. This study aims to design a web scraping framework for acquiring textual data from Platform X to support research on adolescent sexual behavior deviance. The research adopts a Design Science Research (DSR) approach to develop a structured data acquisition pipeline consisting of requirement analysis, keyword formulation, web scraping design, data harvesting, data preprocessing, and dataset construction. The proposed framework emphasizes methodological rigor by integrating data quality assessment, reproducibility, and ethical considerations throughout the data acquisition process. The resulting dataset comprises structured textual data and relevant metadata that are prepared for subsequent analytical stages, such as natural language processing, text mining, machine learning, and behavioral pattern analysis. Furthermore, the proposed web scraping framework provides a systematic approach for researchers to collect social media data efficiently while ensuring data consistency and traceability. This study contributes to the field of social media analytics by providing a replicable methodology for acquiring textual data from Platform X and establishing a reliable foundation for future research on adolescent online behavior and digital risk assessment. The findings are expected to facilitate the development of evidence-based analytical models for understanding behavioral patterns in digital environments while supporting future studies employing Big Data Analytics and Computational Social Science.

Keywords: Web Scraping, Platform X, Data Acquisition, Social Media Analytics, Text Mining, Big Data, Adolescent Behavior, Design Science Research, Natural Language Processing

Corresponding Author:
Yanti Yusman
Faculty Sain Komputasi dan Kecerdasan Digital , Program Study Sistem Informasi ,Universitas Pembangunan Panca Budi, Medan, Indonesia
Email : yantiyusman@gmail.com

1. INTRODUCTION

The development of information and communication technology has fundamentally changed the way individuals interact, communicate, and build social networks. The digital transformation driven by the development of the internet and social media has created an open, dynamic, and real-time communication space. In this context, social media no longer functions solely as a means of exchanging information but also as a representation of users' behavior, opinions, emotions, and experiences in everyday life. Various social media platforms enable users to generate vast digital footprints, forming a continuously growing data ecosystem. This situation makes social media one of the primary data sources in research based on Big Data Analytics, Computational Social Science (CSS), and Natural Language Processing (NLP) to understand various social phenomena more comprehensively [1]

Among social media user groups, adolescents have the highest level of digital activity. As digital natives, adolescents utilize social media not only for communication but also as a means to express their identity, build social relationships, obtain information, and discuss various personal and sensitive issues. The high intensity of social media use results in various forms of adolescent interactions being recorded in digital data that can be used to understand the dynamics of their behavior in cyberspace. Various studies have shown that social media has become a forum for discussions on mental health, interpersonal relationships, identity exploration, and even sexual behavior, which were previously difficult to observe through conventional research approaches due to social stigma and limited access to respondents [2]

This phenomenon presents new opportunities for researchers to utilize social media data as a more naturally occurring source of information compared to surveys or interviews, which are often influenced by social desirability bias. By analyzing data spontaneously generated by users, researchers can obtain a more objective picture of communication patterns, behavioral tendencies, and changes in discourse developing in society. In the context of research on adolescent sexual deviant behavior, a digital data-based approach becomes increasingly important given the sensitive nature of the topic, which makes respondents less likely to openly disclose their experiences or views in traditional research methods. Therefore, analyzing social media text data can be a methodological alternative capable of generating richer information about adolescent digital behavior [3]

One platform that exhibits these characteristics is Platform X, which allows users to express their opinions, experiences, and responses to various issues through real-time text uploads. Unlike visual-based platforms, Platform X generates textual data that is relatively easy to analyze using various computational approaches such as text mining, sentiment analysis, topic modeling, and machine learning. Furthermore, the conversational structure formed through uploads, replies, quotes, and reposts provides information about the dynamics of social interactions that can be used to identify communication patterns on specific issues. These characteristics make Platform X a relevant data source for Big Data-based social behavior research.

However, utilizing social media data as a research source is not without various methodological challenges. One of the main challenges is obtaining relevant, high-quality, and scientifically accountable data. Unlike survey-based research, which has relatively standardized data collection procedures, research using social media requires a systematically designed data acquisition process to produce a dataset representative of the phenomenon being studied. Mistakes in determining data collection strategies, keyword selection, data collection timeframes, and storage processes can create biases that impact the validity of subsequent analyses. Therefore, the quality of Big Data-based research is determined not only by the analytical methods used but also by the quality of the data acquisition process [4]

In recent years, web scraping has become a widely used approach in social media-based research to automatically acquire public data. Web scraping enables systematic data retrieval by utilizing algorithms capable of extracting information from web pages and social media platforms according to predetermined parameters. Compared with manual data collection processes, web scraping offers greater efficiency in large-scale data acquisition while supporting continuous data updating. However, implementing web scraping requires careful design to ensure the data collection process adheres to the principles of validity, reproducibility, and research ethics [5]

Previous research has generally focused on developing analytical models, such as text classification, sentiment detection, and behavioral prediction using machine learning algorithms. Conversely, research specifically addressing how to design a data acquisition process through web scraping is relatively limited. This process is a fundamental step that determines the quality of the dataset used in all subsequent analysis. Datasets created without a clear acquisition procedure have the potential to produce inconsistent data, be difficult to replicate, and have low reliability. Thus, a research gap exists, highlighting the need to develop a systematic and well-documented data acquisition framework.

In addition to technical aspects, research on adolescent sexual deviant behavior also requires attention to research ethics. Data obtained from social media, even though public, still has the potential to contain sensitive information. Therefore, the data acquisition process must adhere to the principles of privacy protection, data anonymization, methodological transparency, and research reproducibility. [6]emphasized that social media-based research should not rely solely on computational performance evaluations but should also integrate a human-centered approach through the involvement of experts in determining analysis categories and evaluating model interpretation results. This approach is especially important in research related to human behavior, as interpretation of text data cannot always be optimally achieved through automated approaches alone.

Based on the description, this study aims to design a systematic web scraping method as a mechanism for acquiring text data on Platform X to support research on adolescent sexual deviant behavior. Unlike previous studies that focused more on the data analysis stage, this study focuses on developing data acquisition stages that include designing data search strategies, compiling keywords, data retrieval, dataset storage, process documentation, and data preprocessing stages to produce datasets that have high quality, consistency, and reproducibility. The results of this study are expected to provide methodological contributions to the development of Big Data-based research, especially in the fields of Information Systems, Social Media Analysis, and Computational Social Science, while providing a strong foundation for the development of digital behavior analysis models in subsequent research stages.

2. RESEARCH METHODOLOGY

This study adopts the Design Science Research (DSR) approach as the primary methodological framework in designing a web scraping process for text data acquisition on Platform X. The DSR approach was chosen because this study is oriented towards developing artifacts in the form of systematic, documented, and replicable data acquisition process designs to support research on adolescent sexual deviant behavior. In contrast to explanatory research that focuses on hypothesis testing, Design Science Research emphasizes the process of designing solutions to a problem through the development of artifacts that have both scientific contributions and practical value [7] (Peffers et al., 2007; Hevner et al., 2004). In the context of this study, the resulting artifact is a web scraping pipeline that includes the process of identifying data needs, developing data retrieval strategies, data acquisition, dataset storage, data pre-processing, and process documentation that supports research reproducibility.

The use of the Design Science Research approach is also based on the need to produce a data acquisition design that is not only technically effective but also meets the principles of validity, transparency, and reproducibility. Big Data-based research relies not only on analysis algorithms but is also heavily influenced by the quality of the data obtained during the acquisition process. Therefore, the web scraping process design must be able to produce consistent, documented, and high-quality data to support further analysis using Natural Language Processing, Machine Learning, and Computational Social Science techniques. Furthermore, all research stages are designed with consideration of the ethical principles of digital research through the use of publicly available data and the implementation of data storage mechanisms that maintain the security and confidentiality of user information [8].

2.1 Research Stages

The research stages are arranged systematically to ensure that the data acquisition process takes place in a structured, documented and replicable manner. The first stage begins with identifying research needs (problem identification) through a literature review regarding adolescent sexual deviant behavior, the characteristics of social media as a source of research data, as well as various web scraping approaches that have been used in previous research. This stage aims to determine the scope of the research, the characteristics of the data required, and the variables that will be used as a basis for preparing a data collection strategy.

The next stage is a data requirements analysis focused on determining the types of data to be collected from Platform X. At this stage, data attributes deemed relevant to the research objectives are identified, such as tweet content, timestamp, language, number of replies, reposts, and likes. These attributes are determined based on the need for digital behavior analysis, which considers not only message content but also the temporal context and level of user interaction. [9] explain that data containing temporal information allows researchers to conduct longitudinal analysis of topic changes and emotional dynamics within a digital community.

Next, a web scraping strategy was designed, including keyword development, search operators, data retrieval timeframes, data filtering mechanisms, and raw data storage design. The search strategy was based on a synthesis of literature on adolescent sexual behavior, reproductive health, and terms commonly used in digital communication, resulting in data relevant to the research objectives. Keyword design was carried out iteratively through an evaluation of initial search results to increase the relevance of the data obtained.

The next stage is the data acquisition process using a web scraping mechanism designed to automatically retrieve public data from Platform X. All successfully obtained data is stored in the form of a raw dataset that still maintains the original structure of the data collection results. Raw data storage is carried out as part of the research reproducibility strategy to allow for validation processes and data re-retrieval if necessary. [10]emphasizes that the separation between raw data and processed data is one of the main principles in improving the reproducibility of data-driven research.

The final stages include data pre-processing, documentation of the research process, evaluation of dataset quality, and compilation of a data dictionary describing all attributes contained in the dataset. All stages are systematically

DOI: https://doi.org/10.64803/jocsaic.v3i1.182

documented, resulting in a transparent research pipeline that can be easily replicated by other researchers in similar research contexts.

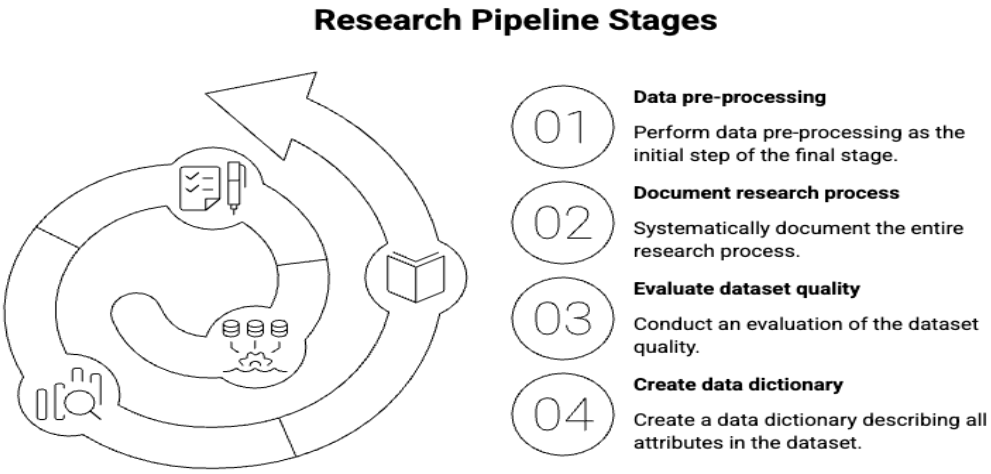

Figure 1. Research Pipeline Stages

2.2 Research Framework

The research framework developed in this study integrates the concepts of Big Data Analytics, Web Scraping, Natural Language Processing, and Design Science Research principles into a single, systematic workflow. The research framework begins with identifying information needs based on the research objectives, followed by developing a data collection strategy through keyword selection and search parameters. Next, data acquisition using web scraping results in a raw textual dataset that still requires validation and pre-processing.

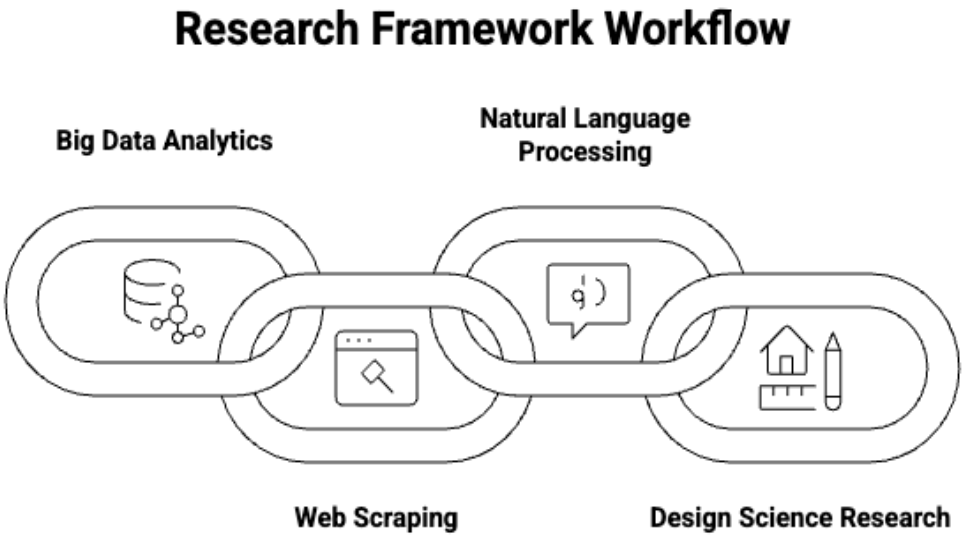

Figure 2 . Research Framework Workflow

After the data was successfully collected, a data cleaning process was carried out by removing duplicate data, URLs, emojis, special characters, and text that lacked relevant information. The next stage was the preparation of a structured dataset that met data quality standards, ready for use in further research, such as sentiment analysis, topic modeling, text classification, and behavioral pattern detection using a Machine Learning approach. The resulting framework not only describes the technical flow of data collection but also emphasizes the importance of process documentation, metadata recording, and dataset management that supports research reproducibility, as recommended by [10], [11].

2.3 Web Scraping Architecture

The web scraping architecture developed in this study uses a modular approach so that each process can be managed and evaluated independently. The architecture consists of five main components: the Query Builder, Scraping Engine, Raw Data Repository, Preprocessing Module, and Structured Dataset Repository.

The Query Builder module is responsible for compiling search parameters based on keywords, time range, and language. These parameters are then forwarded to the Scraping Engine, which automatically retrieves data from Platform X. The retrieved data is then stored in the Raw Data Repository as the primary archive before the cleaning process. Storing raw data is crucial for maintaining research transparency, as it allows for retrieval of the entire data processing process if necessary.

The next stage is carried out by the Preprocessing Module, which is responsible for text normalization, duplication removal, URL removal, emoji removal, symbols removal, user accounts removal, and characters without analytical value. The preprocessed dataset is then stored in a Structured Dataset Repository, ready for use in the next analysis stage. This modular approach is widely used in Big Data research because it increases system flexibility while simplifying research maintenance and reproduction [12], [13]

2.4 Data Acquisition Process

The data acquisition process was carried out systematically using a keyword-based search strategy developed based on a literature review on adolescent sexual behavior, reproductive health, and terms frequently used in digital communication. The keywords used included not only formal terms but also variations of everyday language (slang) (Placeholder2)commonly used by social media users, thereby increasing the scope of the data collected.

During the acquisition process, each piece of data collected is accompanied by supporting attributes such as upload identity, publication time, text content, number of interactions, and language used. All of this metadata is stored alongside the primary data to support temporal-based digital behavior analysis and interaction network analysis in subsequent research. Furthermore, the entire data collection process is documented through a system log that records the time of data collection, search parameters, the amount of data collected, and the status of the scraping process, thus supporting the audit trail and research reproducibility [14]

2.5 Data Preprocessing

The pre-processing stage aims to improve data quality before it is used in further analysis. This process includes removing duplicate data, removing URLs, emojis, special characters, hashtags, and user accounts, and normalizing letters to a uniform format. Furthermore, incomplete data and posts irrelevant to the research objectives are identified so they can be eliminated from the final dataset.

Pre-processing also included compiling a data dictionary, validating the dataset structure, and evaluating data distribution based on time and language characteristics. All pre-processing stages were documented in detail to allow for retracement of the data transformation process. This approach aligns with the principle of methodological reproducibility, which emphasizes the importance of documenting each stage of data processing so that research can be consistently replicated by other researchers [4] Furthermore, this study adopted the principle of human-centered data processing, incorporating manual validation of data cleaning results to ensure that the retained information still represents the context of the behavior being studied [15]

3. RESULT AND DISCUSSION

This section presents the implementation results of the web scraping design developed to obtain text data from Platform X as a research data source regarding adolescent sexual deviant behavior. The main focus of the discussion is not only on the success of the data acquisition process, but also on the evaluation of the system design, the quality of the resulting dataset, and the suitability of the data acquisition pipeline to the needs of Big Data Analytics-based research. The results obtained are analyzed in stages starting from the implementation of the web scraping architecture, the data retrieval process, the characteristics of the successfully collected dataset, and the evaluation of data quality after going through the pre-processing stage. This approach is carried out to ensure that the resulting dataset has a consistent structure, is well-documented, and meets the principles of reproducibility, so that it can be used as a basis for developing further analysis using Natural Language Processing (NLP), Machine Learning, and Computational Social Science (CSS) techniques.

In addition to evaluating the technical aspects, the discussion in this section also highlights the methodological contributions of the proposed web scraping pipeline. The system's effectiveness is assessed based on its ability to automate the public data acquisition process, maintain data integrity through raw dataset storage, and produce a structured dataset ready for use in subsequent analysis stages. Therefore, this section not only presents the results of the system implementation but also discusses the advantages, limitations, and implications of using web scraping as a data acquisition method in digital behavior research on social media. This approach aligns with the principles of Design Science Research, which emphasize that a research artifact should be evaluated based on its function, quality, and contribution to solving the research problem [16], [17]

3.1 Web Scraping Design

After the web scraping design process is completed, the next step is to acquire text data from Platform X based on predetermined search parameters. The scraping process is carried out automatically using keywords compiled based on a literature review on adolescent sexual deviant behavior and various terms commonly used in digital communication. All successfully obtained data is stored in raw dataset form, which still maintains the original structure of the acquisition results so that it can be used as a basis for subsequent pre-processing and analysis. Storing raw data is also part of the application of the principles of data provenance and research reproducibility, namely ensuring that each stage of data transformation can be traced back throughout the research process [18]

As an illustration of the results of implementing web scraping, Table 1 displays a small portion of the data that was successfully obtained during the acquisition process. This table only presents 10 sample data as a representation of the structure of the resulting dataset, while the entire research dataset consists of 500 text data collected from Platform X during the data collection period. Each data contains main information in the form of Tweet ID, publication time (timestamp), user name, text content (tweet content), language, number of replies, number of reposts and number of likes. These attributes provide textual information as well as metadata needed to support digital behavior analysis at the next stage.

Table 1. Sample of the Raw Dataset Collected from Platform

Tweet ID	Timestamp	User name	Tweet Content	Language	Reply	Repost	Likes
19452001345872 1	2025-01-05 09:15:24	user001	Pendidikan kesehatan reproduksi perlu diberikan sejak usia remaja agar mereka memahami risiko perilaku seksual.	id	5	18	72
19452001345872 2	2025-01-06 11:42:08	user002	Banyak remaja memperoleh informasi mengenai hubungan seksual dari media sosial dibandingkan sekolah.	id	8	25	96
19452001345872 3	2025-01-07 14:05:33	user003	Orang tua perlu lebih aktif mengawasi penggunaan media sosial oleh anak-anak mereka.	id	3	12	54
19452001345872 4	2025-01-08 18:20:47	user004	Edukasi digital dapat membantu mengurangi penyebaran konten yang berisiko bagi remaja.	id	4	15	63
19452001345872 5	2025-01-09 20:16:59	user005	Banyak konten di internet yang memengaruhi cara remaja memandang hubungan dan seksualitas.	id	7	20	81
19452001345872 6	2025-01-10 08:31:16	user006	Penting bagi sekolah untuk memberikan pendidikan karakter dan literasi digital secara berkelanjutan	id	2	9	40
19452001345872 7	2025-01-11 13:10:44	user007	Orang tua perlu lebih aktif mengawasi penggunaan media sosial oleh anak-anak mereka.	id	6	17	68
19452001345872 8	2025-01-12 16:54:27	user008	Penggunaan media sosial tanpa pengawasan dapat meningkatkan paparan terhadap konten yang tidak sesuai usia.	id	9	28	104
19452001345872 9	2025-01-13 21:38:55	user009	Literasi digital merupakan salah satu upaya untuk mencegah perilaku berisiko pada remaja.	id	4	13	59

1945200 1345873 0	2025-01-14 10:25:19	user0 10	Kolaborasi antara keluarga, sekolah, dan pemerintah diperlukan untuk menciptakan lingkungan digital yang lebih aman.	id	5	16	75

Based on Table 1, it can be observed that each successfully acquired data has a consistent attribute structure, thus facilitating the data processing process in the next stage. In addition to containing uploaded text, the dataset also stores various metadata that supports temporal analysis and user interaction levels, such as publication time, number of replies, number of reposts, and number of likes. The existence of this metadata allows researchers not only to analyze the text content, but also to evaluate the pattern of information dissemination and the dynamics of user interaction on Platform X. Thus, the obtained dataset has met the basic requirements for research based on Big Data Analytics, Natural Language Processing (NLP), and Computational Social Science (CSS).

3.2 Data Acquisition Results

The web scraping implementation successfully built a data acquisition mechanism capable of collecting public text data from Platform X according to predetermined search parameters. Based on the data retrieval process, the system successfully obtained 500 text data stored in CSV format. Each data consists of several main attributes, namely tweet identifier, timestamp, username, tweet content, language, reply count, repost count, and like count. This data structure provides textual information as well as metadata that can be used to support a more comprehensive analysis of digital behavior.

The success of the scraping process demonstrates that the designed pipeline is capable of automating the data collection process, which was previously performed manually. All data was collected consistently with a uniform attribute structure, facilitating subsequent analysis. Furthermore, storing metadata such as publication date allows future research to conduct temporal analysis to identify changes in discussion patterns over time, as described by [3]

3.3 Dataset Characteristics

After the acquisition process was complete, the dataset characteristics were evaluated to ensure data quality before being used in the next analysis stage. The initial dataset obtained through scraping still contained various common characteristics of social media data, such as the presence of URLs, emojis, symbols, special characters, variations in upper and lower case, and the possibility of duplicate records. Therefore, a preprocessing stage was performed, which included text cleaning, case normalization, URL removal, special character removal, and duplicate data elimination.

The pre-processing stage produces a more consistent and higher-quality dataset than the raw dataset. The normalization process reduces spelling variations, thereby reducing the complexity of the text corpus, while the removal of duplicate data reduces the potential for bias that could influence the analysis results. Furthermore, all attributes contained in the dataset are documented in a data dictionary so that each variable has a clear definition and can be reused in subsequent research. This documentation is part of the implementation of the principle of reproducible research, which ensures that the entire data processing process can be systematically traced [4]

3.4 Discussion

The research results show that the web scraping approach can be an effective alternative for acquiring social media data for digital behavior research. Compared to manual data collection methods, web scraping allows for automated, fast, and consistent data acquisition, resulting in large datasets with a relatively low error rate. This capability is crucial in Big Data-based research, as the quality of analysis is significantly influenced by the completeness and consistency of the collected data.

From a methodological perspective, this study demonstrates that the success of the scraping process is determined not only by the software used, but also by the keyword development strategy and search parameters. Keywords developed based on a literature review yield more relevant data than those using general keywords. Therefore, the Requirements Analysis and Keyword Engineering processes are crucial steps in contributing to the quality of the resulting dataset.

In addition, implementing a modular architecture provides high flexibility in system development. Each module can be modified without changing the entire pipeline, making it easier to adapt to changes in policies or data access mechanisms on Platform These results support the research of [5], which explains that API-based or web scraping data acquisition systems should be designed using a modular architecture to make them easier to maintain and develop.

This study also demonstrates the importance of reproducibility in Big Data-based research. The entire process, from keyword development and search parameters, data retrieval, raw dataset storage, and pre-processing, is systematically documented through research logs and data dictionaries. This approach allows for repeatable research at different time periods using the same procedures, thereby enhancing the validity of the research methodology. This finding aligns with [10]recommendation, which emphasizes that documentation of data acquisition and processing is a key component in improving the quality of data-driven research.

However, this study has several limitations. The data obtained only came from public posts, so it does not represent the overall activity of Platform X users. Furthermore, changes in data access policies or search mechanisms on the platform may affect the amount of data collected. Therefore, future research can develop data acquisition mechanisms that integrate various social media platforms and utilize Natural Language Processing and Machine Learning approaches to perform a more in-depth classification of digital behavior.

4. CONCLUSIONS

This research successfully designed and implemented a web scraping mechanism to acquire text data from Platform X as a data source in research on adolescent sexual deviant behavior. The developed pipeline consists of several integrated stages, namely problem identification, requirements analysis, keyword preparation, query generation, web scraping, raw data storage, data preprocessing, dataset evaluation, and the formation of a structured dataset. This approach allows the data acquisition process to be carried out systematically, documented, and meets the principle of reproducibility, so that each stage of the research can be traced and replicated in subsequent research. The implementation results show that the web scraping design is able to automatically obtain public data from Platform The existence of these attributes produces a dataset that not only supports text content analysis, but also allows temporal analysis and user interaction analysis. In addition, the implementation of the pre-processing stage succeeded in improving the quality of the dataset through the process of cleaning text, normalizing, removing duplicate data, and documenting attributes in the form of a data dictionary, so that the dataset became more consistent and ready to be used in advanced analysis stages. The main contribution of this research lies in the development of a web scraping pipeline that can be used as a framework in research based on Big Data Analytics, Natural Language Processing (NLP), and Computational Social Science (CSS). The proposed modular architecture provides flexibility for development on various social media research topics, while simplifying the system's maintenance, development, and evaluation processes. Thus, this research not only produces a structured dataset but also provides a methodological contribution to the development of transparent, efficient, and replicable digital data acquisition mechanisms. However, this study has several limitations. Data acquisition was conducted solely on publicly available data on Platform X, thus not representing all forms of user interaction on social media. Furthermore, changes in platform policies, search mechanisms, and data access restrictions can impact the amount and characteristics of data collected. Therefore, further research is recommended to develop data acquisition mechanisms that support the integration of multiple social media platforms, implement more sophisticated natural language processing techniques, and develop machine learning- and deep learning-based analytical models to more comprehensively identify and classify digital behavior.

REFERENCES

[1] D. M. J. Lazer *et al.*, “Computational social science: Obstacles and opportunities,” *Science (1979).*, vol. 369, no. 6507, pp. 1060–1062, 2020.

[2] A. S. Walsh, “Representations of Gender-Diverse Characters in Adventure Time and Steven Universe,” *Stud. Media Commun.*, vol. 8, no. 2, pp. 61–69, 2020, doi: 10.11114/smc.v8i2.5048.

[3] M. Stevenson, “The value of text for small business default prediction: A Deep Learning approach,” *Eur. J. Oper. Res.*, vol. 295, no. 2, pp. 758–771, 2021, doi: 10.1016/j.ejor.2021.03.008.

[4] S. S. Kunia and S. S. Othman, “Investigative reporting pattern of tempo weekly news magazine,” *Humanities and Social Sciences Reviews*, vol. 7, no. 1, pp. 19–30, 2019, doi: 10.18510/hssr.2019.713.

[5] H. E. Kazdouh, A. El-Ammari, S. Bouftini, S. E. Fakir, and Y. E. Achhab, “Perceptions and Intervention Preferences of Moroccan Adolescents, Parents, and Teachers Regarding Risks and Protective Factors for Risky Sexual Behaviors Leading to Sexually Transmitted Infections in Adolescents: Qualitative Findings,” *Reprod. Health*, vol. 16, no. 1, 2019, doi: 10.1186/s12978-019-0801-y.

[6] M. Kassab, B. Amaba, E. S. Pound, and B. Fox, “Is the Solution for Cybercrime Also a Path to Greater Productivity?,” *Computer (Long. Beach. Calif).*, vol. 56, no. 11, pp. 91–94, 2023, doi: 10.1109/MC.2023.3290263.

[7] M. Shuhufi, M. Qadaruddin, J. Basyir, M. M. Yunus, and N. M. Nur, “Islamic Law and Social Media: Analyzing the Fatwa of Indonesian Ulama Council Regarding Interaction on Digital Platforms,” *Samarah*, vol. 6, no. 2, pp. 823–843, 2022, doi: 10.22373/sjhk.v6i2.15011.

[8] J. F. Kusuma and A. Chowanda, "Indonesian Hate Speech Detection Using IndoBERTweet and BiLSTM on Twitter," *International Journal on Informatics Visualization*, vol. 7, no. 3, pp. 773–780, 2023, doi: 10.30630/joiv.7.3.1035.

[9] N. Kerekes *et al.*, "Changes in adolescents' psychosocial functioning and well-being as a consequence of long-term covid-19 restrictions," *Int. J. Environ. Res. Public Health*, vol. 18, no. 16, 2021, doi: 10.3390/ijerph18168755.

[10] U. Alter, C. Dang, Z. J. Kunicki, and A. Counsell, "The VSSL scale: A brief instructor tool for assessing students' perceived value of software to learning statistics," *Teach. Stat.*, vol. 46, no. 3, pp. 152–163, 2024, doi: 10.1111/test.12374.

[11] X. Guzmán-Guzmán, E. R. Núñez-Valdez, R. Vásquez-Reynoso, A. Asencio, and V. García-Díaz, "SWQL: A new domain-specific language for mining the social Web," *Sci. Comput. Program.*, vol. 207, Jul. 2021, doi: 10.1016/j.scico.2021.102642.

[12] A. Arfan, A. R. A. Rizal, and M. W. Kawuryan, "Value-Belief Communication in Social Media: An Analysis of the Influencers' Effect on TikTok," *Jurnal Komunikasi: Malaysian Journal of Communication*, vol. 39, no. 4, pp. 428–444, 2023, doi: 10.17576/JKMJC-2023-3904-23.

[13] M. Birjali, A. Beni-Hssane, and M. Erritali, "Analyzing Social Media through Big Data using InfoSphere BigInsights and Apache Flume," in *Procedia Computer Science*, E. Shakshuki, Ed., Elsevier B.V., 2017, pp. 280–285. doi: 10.1016/j.procs.2017.08.299.

[14] C. S. Andreassen *et al.*, "The Relationship Between Addictive Use of Social Media and Video Games and Symptoms of Psychiatric Disorders: A Large-Scale Cross-Sectional Study.," *Psychology of Addictive Behaviors*, vol. 30, no. 2, pp. 252–262, 2016, doi: 10.1037/adb0000160.

[15] S. M. Doornwaard, T. t. Bogt, E. Reitz, and R. J. J. M. van den Eijnden, "Sex-Related Online Behaviors, Perceived Peer Norms and Adolescents' Experience With Sexual Behavior: Testing an Integrative Model," *PLoS One*, vol. 10, no. 6, p. e0127787, 2015, doi: 10.1371/journal.pone.0127787.

[16] J. K. Haner and R. K. Knake, "Breaking botnets: A quantitative analysis of individual, technical, isolationist, and multilateral approaches to cybersecurity," *J. Cybersecur.*, vol. 7, no. 1, 2021, doi: 10.1093/cybsec/tyab003.

[17] J.-C. Plantin and A. Punathambekar, "Digital media infrastructures: pipes, platforms, and politics," *Media Cult. Soc.*, vol. 41, no. 2, pp. 163–174, 2019, doi: 10.1177/0163443718818376.

[18] X. Wang, X. Long, G. Li, J. Li, and Y. Zhao, "Application Method and Least Squares Support Vector Machine Analysis of a Heat Pipe Network Leakage Monitoring System Using an Inspection Robot," *Informatica (Slovenia)*, vol. 49, no. 16, pp. 199–212, 2025, doi: 10.31449/inf.v49i16.6990.