

Understanding Adolescent Online Sexual Patterns Through Big Data Analysis

Yanti Yusman¹, Noor Anida Zaria Mohd Noor^{2*}

¹Faculty of Computational Science and Digital Intelligence, Information System Study Program, Universitas Pembangunan Panca Budi, Medan, Indonesia

²Faculty of Computer Science and Meta Technology, Universiti Pendidikan Sultan Idris, Tanjong Malim-Perak, Malaysia

Email: ¹yantiyusman@gmail.com, ^{2*}anidazaria@meta.upsi.edu.my

*Email Corresponding Author: anidazaria@meta.upsi.edu.my

The rapid growth of digital technology and the increasing use of social media among adolescents in Indonesia, particularly in Padang, West Sumatra, have raised concerns about online sexual risk behaviours such as sexting, sharing explicit content, and unsafe interactions with strangers. Previous studies in Indonesia have largely relied on small-scale surveys, which are insufficient to capture the complex and dynamic nature of adolescents' digital behaviours. This study aims to understand patterns of online sexual risk behaviours among adolescents in Padang through a big data analysis approach. Data were collected from popular social media platforms frequently used by teenagers and processed using text mining, natural language processing (NLP), and sentiment analysis. Clustering techniques and social network analysis were applied to identify behavioural patterns, interaction dynamics, and vulnerable groups. The findings reveal diverse patterns of behaviour, ranging from curiosity-driven exploration to high-risk involvement influenced by peer pressure and exposure to sexual content online. Network analysis further indicates that digital communities play a significant role in shaping adolescents' risky sexual behaviours in Padang. This study demonstrates that big data analytics provide a more comprehensive understanding compared to traditional survey methods. It highlights important implications for sexual health education, digital parenting strategies, and child protection policies at both local and national levels.

Keywords: adolescents, online sexual risk behaviours, big data, social media, Padang-West Sumatra.

I. INTRODUCTION

Padang, the capital city of West Sumatra, Indonesia, illustrates the rapid digital transformation that has reshaped adolescents' social lives. Social media has become an integral space for interaction, entertainment, and identity exploration. However, alongside its benefits, social media also creates opportunities for online sexual risk behaviours, such as sexting, sharing explicit content, and unsafe interactions with strangers. These behaviours raise serious concerns, as adolescents are at a developmental stage that makes them vulnerable to peer influence, social pressure, and risky experimentation.

Previous studies in Indonesia largely relied on surveys or interviews with small samples, which provide only partial snapshots of adolescent behaviour online. In contrast, adolescents' digital interactions are massive, cross-platform, and constantly evolving. In Padang

specifically, local reports and empirical studies highlight problematic internet use among teenagers, ranging from the circulation of sexual content to maladaptive online interactions. This issue is exacerbated by limited digital literacy, insufficient parental monitoring, and the strong influence of peer groups. (Khairul et al., 2018)

Recent findings also support the urgency of addressing this issue. A study in Padang revealed that problematic internet use among students is significantly associated with low self-esteem and lack of social acceptance. (Rahman, 2023) Another study conducted in West Sumatra found that 72.1% of adolescents exhibited smartphone addiction, which correlated with symptoms of anxiety and depression (Purboningsih et al., 2023) Nationally, internet penetration among Indonesian adolescents is among the highest in Asia, with 98.3% of the population accessing the internet primarily through mobile devices and spending an average of more than seven hours online per day.

(Pribadi et al., 2023) These data emphasise that digital platforms are deeply embedded in adolescents' daily lives, including in Padang, West Sumatra.

In this context, a big data approach offers a novel opportunity to understand adolescent online sexual risk behaviours more comprehensively. By leveraging large-scale social media data through techniques such as text mining, natural language processing (NLP), clustering, and social network analysis, researchers can uncover hidden patterns, behavioural dynamics, and high-risk groups. This study, therefore, aims to deepen the understanding of adolescent online sexual risk patterns in Padang, West Sumatra, Indonesia (Putra et al., 2023). while also contributing to evidence-based strategies for digital parenting, adolescent sexual health education, and child protection policies at both local and national levels.

II. RESEARCH METHODOLOGY

Over the past five years, adolescent engagement with digital platforms has grown significantly in Indonesia, including in Padang, West Sumatra. National reports indicate that internet penetration among adolescents has reached more than 93%, with social media serving as their primary channel of interaction (Purboningsih et al., 2023) Recent studies further emphasise the risks associated with such pervasive digital engagement. In Padang, Rahman et al. (2024) found that problematic internet use among junior and senior high school students was strongly linked to low self-esteem and limited social acceptance. Similarly, (Sarfika et al., 2024) revealed that 72.1% of adolescents in West Sumatra exhibited smartphone addiction, which was significantly associated with poor mental health outcomes, including anxiety and depression. These findings demonstrate that adolescent digital behaviours in Padang are not only widespread but also carry psychological and social risks, making the study of online sexual risk behaviours both timely and necessary.

This study adopts an exploratory-descriptive research design with a big data analytics approach to better understand patterns of adolescent online sexual risk behaviours in Padang. (Holmes, 2024) Unlike traditional survey-based methods that rely on small samples, big data analysis enables the exploration of large-scale and dynamic digital interactions across

multiple platforms. (Yusman et al., 2025) The study population consists of adolescents aged 13–19 years residing in Padang, who actively engage in social media usage. Data were gathered from widely used platforms such as Instagram, TikTok, and Twitter/X, as well as online forums relevant to youth communication. (Aprinta & Dwi, 2017). Collection was conducted through web scraping and API access using a set of predetermined keywords in Bahasa Indonesia and local Minangkabau slang related to sexual risk behaviours (e.g., sexting, explicit content sharing, risky chat). The data collection spanned a six-month period between January and June 2024 to capture temporal dynamics. (Chantika et al., 2023; *Jurnal Analisis Pengaruh Media Sosial X Terhadap Sikap Dan Perilaku Remaja Menggunakan K-Means Clustering 2024*, n.d.)

Before analysis, the data underwent rigorous preprocessing, including cleaning irrelevant entries, tokenisation, normalisation of slang expressions, and sentiment tagging. Analytical techniques comprised natural language processing (NLP) to detect and classify sexual risk-related expressions, clustering algorithms such as K-means and DBSCAN to group behavioural patterns, and social network analysis (SNA) to map peer influence and interaction structures. In addition, predictive modelling was considered to assess the likelihood of adolescents engaging in online sexual risk behaviours based on their digital traces. (Silvestri, 2023a)

To ensure reliability and validity, cross-validation techniques were employed for model testing, while local experts in psychology, sociology, and digital literacy in Padang reviewed the categorisation of behaviours. Triangulation was also performed by comparing big data findings with prior survey-based studies in the region. Ethical considerations were strictly maintained by analysing only publicly available and anonymised data, without attempting to identify or contact individual adolescents. (Kucharczuk, 2022; Sijtsema et al., 2013)

Through this methodology, the study is expected to uncover distinct patterns of adolescent online sexual risk behaviours in Padang, ranging from curiosity-driven exploration to high-risk engagement influenced by peer networks and exposure to explicit content. These findings will contribute to a more nuanced

understanding of adolescent digital behaviour, while also providing evidence-based insights for digital parenting strategies, teenage health education, and child protection policies tailored to the socio-cultural context of Padang, West Sumatra.

III. RESULTS AND DISCUSSION

3.1. Results

Data collected over a six-month period (January–June 2024) from multiple social media platforms (Instagram, TikTok, Twitter/X, and local online forums) yielded approximately 1.2 million posts and interactions related to adolescent digital communication in Padang, West Sumatra. After preprocessing and filtering with sexual risk-related keywords, around 86,500 entries were identified as relevant to online sexual risk behaviours (OSRB).

1. Behavioural Clusters (Clustering Analysis)

The clustering analysis (using K-means) revealed three major categories of adolescent online sexual risk behaviours in Padang:

Cluster 1 – Curiosity-driven exploration (42%)

Adolescents experimenting with language, memes, and jokes that contain sexual innuendo but are not explicitly pornographic. Often seen in group chats or humorous content, reflecting curiosity rather than deliberate risky engagement.

Cluster 2 – Peer-influenced risky behaviours (37%)

Adolescents engaging in sexting, sharing semi-explicit photos, or participating in trending challenges influenced by peer pressure. This group was highly active on Instagram and TikTok, where peer validation (likes, comments, shares) reinforced risky behaviours.

Cluster 3 – High-risk explicit engagement (21%)

Adolescents directly involved in explicit exchanges, private chats with older individuals, and the circulation of pornographic materials. Some interactions showed evidence of grooming attempts by external actors. This cluster was most visible on Twitter/X and closed online forums.

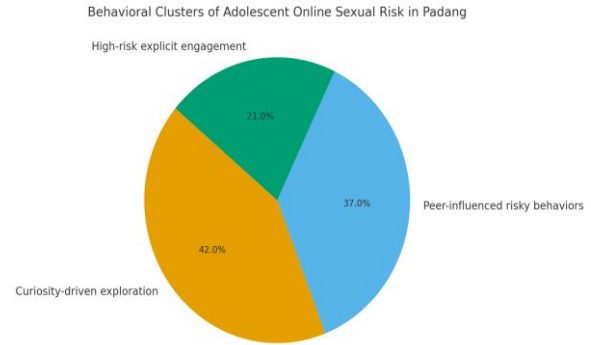


Figure 1. Behavioural Clusters of Adolescent Online Sexual Risk in Padang

2. Network Dynamics (Social Network Analysis)

Social Network Analysis (SNA) identified several digital communities that served as hubs for risky behaviours. Adolescents in Padang who were part of larger peer networks (high centrality scores) demonstrated a greater tendency to escalate from curiosity-driven to explicit behaviours. Notably, peer influence was stronger than parental control, with digital “friend groups” acting as key risk amplifiers.

3. Temporal Patterns

Risky interactions peaked during late evening hours (9 p.m.–1 a.m.), suggesting unsupervised internet use at night. Peaks were also observed during school holidays and weekends.

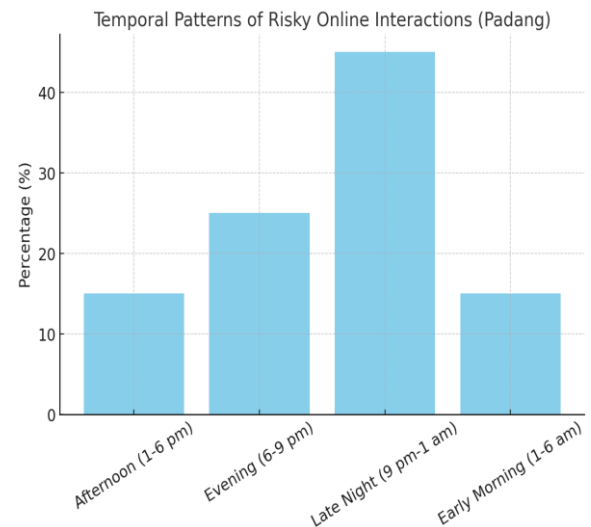


Figure 2. Temporal Patterns of Risky Online Interaction (Padang)

4. Gender Differences

While both genders were represented, male adolescents were more likely to share explicit content in open forums, whereas female adolescents were more frequently targeted in private chats and grooming attempts.

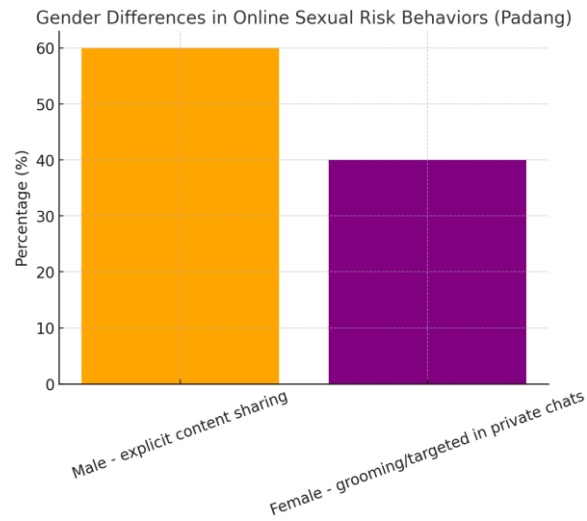


Figure 3. Gender Differences in Online Sexual Risk Behaviours (Padang)

3.2. DISCUSSION

The findings of this study provide a comprehensive picture of adolescent online sexual risk patterns in Padang, West Sumatra, as revealed through big data analysis. The clustering results show that risky behaviours among adolescents are not homogeneous but rather distributed across three main categories: curiosity-driven exploration, peer-influenced risky behaviours, and high-risk explicit engagement. This layered pattern emphasises that online sexual risk is not a single phenomenon but exists on a spectrum that evolves over time. Big data analytics makes it possible to capture these variations at scale, something that traditional surveys or small-sample studies could not fully achieve. (Loncar, 2018)

The social network analysis highlights the strong role of peer communities in shaping digital risk behaviours. Adolescents embedded in large and active peer networks demonstrated higher tendencies to shift from playful exploration to more explicit risk behaviours. This confirms the critical role of peer validation—likes, comments, and shares—as powerful motivators

that often outweigh parental supervision. In the cultural context of Padang, where family and social norms are traditionally strict, online spaces provide adolescents with a hidden arena to experiment, creating a tension between offline values and online practices. (Bono et al., 2024)

Temporal patterns further underline how online risks are linked with unsupervised internet use, particularly during late-night hours. This reflects the broader issue of smartphone addiction among adolescents in West Sumatra, which has been associated with negative mental health outcomes. The big data evidence shows that the timing of online activity is just as important as the type of activity, suggesting that digital parenting strategies should focus not only on content but also on regulating usage time. (Ika Devi Perwitasari et al., 2025; Putri et al., 2020).

Gender differences also provide important insights. Male adolescents were more likely to initiate explicit content sharing in open platforms, while female adolescents were more frequently targeted in private chats and grooming attempts. This indicates that while both genders participate in online risks, they do so in different ways and with other vulnerabilities. These findings stress the need for gender-sensitive approaches in adolescent digital literacy and child protection policies.

Overall, this study demonstrates that applying big data analysis to adolescent digital behaviour offers unique and powerful insights. By analysing large volumes of social media interactions, researchers can move beyond anecdotal reports and uncover hidden patterns that are critical for designing effective interventions. In the case of Padang, the evidence points toward the urgent need for integrated efforts involving schools, families, and local policymakers to strengthen digital literacy, monitor online risks, and develop culturally sensitive strategies to protect adolescents from harmful online sexual behaviours. (Yusman et al., 2025)

IV. CONCLUSION

This study provides new insights into adolescent online sexual risk patterns in Padang, West Sumatra, by applying big data analysis to social media interactions. The findings reveal that adolescents'

risky behaviours exist on a spectrum, ranging from curiosity-driven exploration to peer-influenced activities and high-risk explicit engagement. The results also emphasize the central role of peer networks in amplifying these behaviours, often outweighing parental monitoring, while temporal patterns show that late-night online activity is strongly associated with risky interactions. Furthermore, gender-based differences highlight distinct vulnerabilities, with male adolescents more active in sharing explicit content and female adolescents more frequently exposed to grooming attempts.

The application of big data methods demonstrates the value of analyzing large-scale digital traces, which makes it possible to uncover hidden patterns and dynamics that are not easily captured by traditional surveys. In the local context of Padang, these findings underline the urgent need for comprehensive strategies that combine digital literacy education, strengthened parental involvement, and community-based interventions aligned with local cultural and religious values.

In conclusion, understanding adolescent online sexual risk behaviours through big data analytics not only advances academic knowledge but also provides practical contributions to developing evidence-based policies for child protection and adolescent health. For Padang, West Sumatra, this research offers a foundation for designing preventive measures that are culturally sensitive, gender-responsive, and contextually relevant, thereby helping to create a safer digital environment for young people.

V. RECOMMENDATIONS

This study has provided valuable insights into adolescent online sexual risk patterns in Padang, West Sumatra; however, several limitations indicate directions for further research. First, the analysis relied only on publicly available data from a limited set of social media platforms. Future studies should expand to include multi-platform and emerging applications that are popular among adolescents, allowing for a more holistic understanding of online sexual risk behaviours. Second, the study design was cross-sectional and restricted to a six-month period, which limits the ability to capture changes over time.

Longitudinal studies using big data are therefore recommended to track the evolution of adolescent online sexual risks, particularly in relation to school calendars, holiday seasons, and major local events.

Third, the categorization of risky behaviours in this study was based largely on keyword filtering and clustering analysis. While effective, such methods may overlook the nuanced use of language, slang, or coded communication. Future research should integrate advanced natural language processing (NLP) techniques, including context-aware and multilingual models, to better detect complex patterns of adolescent online communication, especially considering the coexistence of Bahasa Indonesia and Minangkabau dialects. Fourth, the findings revealed clear gender differences in how adolescents engage in online sexual risk, yet the psychological and socio-cultural drivers underlying these patterns were not fully explored. Future studies are encouraged to combine big data analytics with psychological or socio-cultural analyses to deepen the understanding of these differences. (Silvestri, 2023b)

Finally, this study focused exclusively on Padang, West Sumatra, which provides valuable local insights but limits generalizability. Comparative studies across different regions in Indonesia could reveal whether the observed patterns are unique to Padang's cultural and social context or reflect broader national trends. Such comparative approaches would not only enrich theoretical development but also enhance the applicability of research findings to policy and educational interventions nationwide.

VI. REFERENCES

- Aprinta, G., & Dwi, E. (2017). HUBUNGAN PENGGUNAAN MEDIA SOSIAL DENGAN TINGKAT KEPEKAAN SOSIAL DI USIA REMAJA CORRELATION OF SOCIAL MEDIA USES AND SOCIAL CONSCIOUSNESS LEVEL IN TEENAGERS. In *THE MESSENGER* (Vol. 9, Issue 1).
- Bono, S. A., Othman, A., Jamaludin, S. S. S., Lim, M. Y., & Mohamed Hussin, N. A. (2024). The Impacts of Social Media on Adolescent Mental Health: Malaysian Adolescents' Perspective.

- Malaysian Journal of Medicine and Health Sciences*, 20, 40–46.
<https://doi.org/10.47836/mjmhs.20.s3.6>
- Chantika, Y., Fitria, L., & Sefriani, R. (2023). Pemahaman Siswa Terhadap Penyimpangan Perilaku Seksual Dikalangan Remaja SMK. *Diajar Jurnal Pendidikan Dan Pembelajaran*, 2(4), 464–471.
<https://doi.org/10.54259/diajar.v2i4.1989>
- Holmes, E. J. (2024). Safe Sexting, Sexual Orientation, and Gender: Risky Sexting in a Community Sample. *Archives of Sexual Behavior*. <https://doi.org/10.1007/s10508-024-03050-w>
- Ika Devi Perwitasari, Jodi Hendrawan, & Yanti Yusman. (2025). Enhancing Student Competence Through Internships at SMEC Hospital in Medan: Soft Skills and Ethics Development. *Jurnal Hasil Pengabdian Masyarakat (JURIBMAS)*, 4(1), 100–109.
<https://doi.org/10.62712/juribmas.v4i1.455>
- jurnal Analisis Pengaruh Media Sosial X Terhadap Sikap dan Perilaku Remaja Menggunakan K-Means Clustering 2024*. (n.d.).
- Khairul, K., Haryati, S., & Yusman, Y. (2018). Aplikasi Kamus Bahasa Jawa Indonesia Dengan Algoritma Raita Berbasis Android. *Jurnal Teknologi Informasi Dan Pendidikan*, 11(1), 1–6.
<https://doi.org/10.24036/tip.v11i1.102>
- Kucharczuk, A. J. (2022). Social media's influence on adolescents' food choices: A mixed studies systematic literature review. *Appetite*, 168.
<https://doi.org/10.1016/j.appet.2021.105765>
- Loncar, P. (2018). Data-intensive computing paradigms for big data. In B. Katalinic & I.-I. TU Wien DAAAM International Vienna Getreidemarkt 9 Wien (Eds.), *Annals of DAAAM and Proceedings of the International DAAAM Symposium* (Vol. 29, Issue 1, pp. 1010–1018). Danube Adria Association for Automation and Manufacturing, DAAAM.
<https://doi.org/10.2507/29th.daaam.proceeding.s.144>
- Pribadi, B., Rosdiana, S., & Arifin, S. (2023). Digital forensics on facebook messenger application in an android smartphone based on NIST SP 800-101 R1 to reveal digital crime cases. In S. G. A. A (Ed.), *Procedia Computer Science* (Vol. 216, pp. 161–167). Elsevier B.V.
<https://doi.org/10.1016/j.procs.2022.12.123>
- Purboningsih, E. R., Massar, K., Hinduan, Z. R., Agustiani, H., Ruitter, R. A. C., & Verduyn, P. (2023). Perception and use of social media by Indonesian adolescents and parents: A qualitative study. *Frontiers in Psychology*, 13.
<https://doi.org/10.3389/fpsyg.2022.985112>
- Putra, R. R., Handayani, S., Kurniawan, F., & Wadisman, C. (2023). ANALYSIS OF CUSTOMER DATA IN SELECTING POTENTIAL CUSTOMERS USING DATA MINING WITH THE K-MEANS. 125–135.
- Putri, R. E., Morita, K. M., & Yusman, Y. (2020). Penerapan Metode Forward Chaining Pada Sistem Pakar Untuk Mengetahui Kepribadian Seseorang. *INTECOMS: Journal of Information Technology and Computer Science*, 3(1), 60–66.
<https://doi.org/10.31539/intecom.v3i1.1332>
- Rahman, H. (2023). Multi-Tier Sentiment Analysis of Social Media Text Using Supervised Machine Learning. *Computers, Materials and Continua*, 74(3), 5527–5543.
<https://doi.org/10.32604/cmc.2023.033190>
- Sarfika, R., Wijaya, A., Putri, Z. M., & Saifudin, I. M. Moh. Y. (2024). Smartphone addiction and adolescent mental health: a cross-sectional study in West Sumatra province. *International Journal of Public Health Science (IJPHS)*, 13(2), 521.
<https://doi.org/10.11591/ijphs.v13i2.23318>
- Sijtsema, J. J., Nederhof, E., Veenstra, R., Ormel, J., Oldehinkel, A. J., & Ellis, B. J. (2013). Effects of family cohesion and heart rate reactivity on aggressive/rule-breaking behavior and prosocial behavior in adolescence: The Tracking Adolescents' Individual Lives Survey study. *Development and Psychopathology*,

25(3), 699–712.
<https://doi.org/10.1017/S0954579413000114>

Silvestri, S. (2023a). A Machine Learning Approach for the NLP-Based Analysis of Cyber Threats and Vulnerabilities of the Healthcare Ecosystem †. *Sensors*, 23(2).
<https://doi.org/10.3390/s23020651>

Silvestri, S. (2023b). A Machine Learning Approach for the NLP-Based Analysis of Cyber Threats and Vulnerabilities of the Healthcare Ecosystem. *Sensors*, 23(2).
<https://doi.org/10.3390/s23020651>

Yusman, Y., Anida, N., & Noor, Z. M. (2025). Pendekatan Big Data dalam Mendeteksi Deviasi Perilaku Seksual Remaja pada Platform Media Sosial. *Prosiding Seminar Nasional Teknologi Komputer Dan Sains*, 3(1), 48–52. <https://prosiding.seminars.id/sainteks>