



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

FORM PENGAJUAN JUDUL



Nama : Stephani Uli Basa Silitonga

NIM : 201402068

Judul diajukan oleh* : ☐ Dosen
☒ Mahasiswa

Bidang Ilmu (tulis dua bidang) : Data Science and Intelligent System

Uji Kelayakan Judul** : ☐ Diterima ☐ Ditolak

Hasil Uji Kelayakan Judul :

Calon Dosen Pembimbing I: Umayu Ramadhani Putri Nasution S.TI., M.Kom.
(Jika judul dari dosen maka dosen tersebut berhak menjadi pembimbing I)

Calon Dosen Pembimbing II: Sarah Purnamawati S.T., M.Sc.

Medan,

Ka. Laboratorium Penelitian,

* Centang salah satu atau keduanya

** Pilih salah satu

(Fanindia Purnamasari, S.TI., M.IT)

NIP. 198908172019032023



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI


UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

RINGKASAN JUDUL YANG DIAJUKAN

*Semua kolom di bawah ini diisi oleh mahasiswa yang sudah mendapat judul

Judul / Topik Skripsi	IDENTIFIKASI PERNYATAAN MISOGINI BERDASARKAN KOMENTAR MEDIA SOSIAL MENGGUNAKAN BIDIRECTIONAL LONG SHORT-TERM DAN INDOBERT EMBEDDING
Latar Belakang, Penelitian Terdahulu, dan Perbedaan Penelitian	<p>Latar Belakang</p> <p>Di Indonesia, masih sering ditemukan masyarakat yang menganut paham patriarki. Hal ini membuat misogini masih terus berkembang secara luas seiring dengan perkembangan ideologi tersebut. Misogini sendiri memiliki arti kebencian terhadap perempuan dan pemahaman ini telah melahirkan seorang misogynis. Seorang misogynis memiliki sikap yang memandang bahwa perempuan merupakan pihak yang layak untuk disudutkan, ditindas, dan dieksploitasi, serta mendapat kebencian. Selain hal ini, terdapat juga kebiasaan yang meremehkan dan menomorduakan perempuan, menghina dan mendiskriminasi perempuan, serta menganggap bahwa perempuan layak dilecehkan dan diberi kekerasan (Lubis, 2021) (Maharani, 2022). Adapun contoh pernyataan misogini dapat dilihat pada gambar di bawah ini.</p> <div><p>The screenshot shows a social media comment from user @razimansukri5654, posted 2 years ago. The comment is in Indonesian and discusses misogyny and violence against women. It mentions that women are often treated as inferior and are the target of discrimination and violence. The comment also includes a link to a video and a reply button.</p></div> <p>Tentunya, perilaku ini sudah termasuk kedalam tindakan kekerasan terhadap perempuan dan telah melanggar prinsip HAM terkait kesetaraan dan nondiskriminasi. Meskipun demikian, perilaku ini masih tersebar luas dan bahkan dapat ditemukan di media sosial.</p> <p>Penggunaan media sosial yang identik dilakukan secara anonim mengakibatkan sering kali ditemukan pernyataan atau unggahan yang tidak berkenan terhadap sesama masyarakat terutama terhadap perempuan. Hal tersebut diperkuat pernyataannya berdasarkan laporan oleh <i>National Democratic Institute</i> (National Democratic Institute, 2019), dimana dikatakan bahwa perempuan-perempuan di Kolombia, Kenya, dan Indonesia menjadi yang paling umum mengalami kekerasan secara <i>online</i> berupa bahasa yang menghina. Dan dalam Catatan Tahunan (CATAHU) 2023 (Fuad et al., 2023), Komisi Nasional Anti Kekerasan terhadap Perempuan (Komnas Perempuan) mendapatkan pengaduan kasus Kekerasan Siber Berbasis Gender (KSBG) sebanyak 1697 kasus yang dimana 869 kasus diantaranya terjadi di ranah publik. Di antara kasus-kasus tersebut, sebanyak 383 pelaku merupakan teman media sosial korban dan 304 pelaku merupakan orang yang tidak dikenal oleh korban. Selain Komnas Perempuan, terdapat</p>



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

juga catatan pengaduan dari Lembaga Layanan yang menyatakan sebanyak 148 kasus KSBG diantara semua kasus yang diadukan. Banyaknya kasus yang terjadi ini dapat dan telah menimbulkan rasa tidak aman terhadap perempuan dalam menggunakan teknologi seperti halnya media sosial. Bahkan, beberapa perempuan diantaranya menghentikan sementara aktivitas mereka di media sosial dan terdapat juga yang berhenti secara permanen menggunakan media sosial (National Democratic Institute, 2019).

Hal terkait kekerasan melalui jejaring sosial ini tentunya perlu diamati dan dicermati lebih lanjut karena telah merugikan perempuan, terutama terkait misogini yang menjadi salah satu bentuknya. Akan tetapi, untuk membuktikan suatu pernyataan merupakan misogini atau bukan, diperlukan ketelitian dan pemahaman terhadap setiap konteks kalimat yang ada pada komentar. Hal tersebut tentunya akan menghabiskan waktu yang cukup lama untuk membuktikan pernyataannya satu per satu. Sehingga diperlukan suatu pendekatan yang dapat mempermudah proses tersebut.

Pendekatan yang dapat dilakukan ialah dengan melakukan identifikasi pernyataan misogini menggunakan metode pembelajaran mesin. Sebelumnya telah dilakukan beberapa penelitian yang serupa, diantaranya penelitian yang berjudul “*Identifikasi Pernyataan Misogini Berbahasa Indonesia Berdasarkan Komentar Youtube Menggunakan GloVe Embedding dan Random Forest Classifier*” (Damanik, 2021). Penelitian tersebut berupa skripsi yang melakukan identifikasi pernyataan misogini yang diinput dalam bentuk dokumen menggunakan algoritma *Random Forest (RF)* dan *GloVe Embedding* dengan hasil akurasi 92,5%. Penelitian lainnya berjudul “*Misogyny Speech Detection Using Long Short-Term Memory and BERT Embeddings*” (Angeline et al., 2022). Penelitian ini melakukan pendeteksian pernyataan misogini pada Twitter dengan menggunakan *Logistic Regression (LR)*, *Convolutional Neural Network (CNN)*, dan *Long Short-Term (LSTM)*, serta TF-IDF, Keras, dan BERT sebagai *embedding*. Hasil penelitian ini mendapatkan akurasi terbaik sebesar 86,15% melalui LSTM dengan bantuan BERT. Terdapat juga penelitian yang berjudul “*Hateminers: Detecting Hate Speech Against Women*” (Saha et al., 2018). Penelitian tersebut melakukan deteksi (*task A*) dan klasifikasi (*task B*) pernyataan misogini, dimana klasifikasi tersebut dilakukan berdasarkan lima kategori misogini dan juga berdasarkan target misogini antara individu atau kelompok. Penelitian dilakukan dengan menggunakan *Logistic Regression (LR)*, *XGBoost (XGB)*, dan *CatBoost*, serta *Sentence Embeddings*, TF-IDF, dan *Bag of Words (BoW)* sebagai ekstraksi fitur. Hasil dari penelitian ini menyatakan bahwa akurasi terbaik untuk *task A* diraih oleh *Logistic Regression* sebesar 70,4% dan *task B* diraih oleh *Catboost* dengan nilai rata-rata F-Measure 37%.



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

Selanjutnya, terdapat penelitian yang berjudul “*Misogynous Text Classification Using SVM and LSTM*” (Devi & Saharia, 2020). Penelitian ini melakukan klasifikasi pernyataan termasuk misogini atau bukan dalam dataset berbahasa India dan Inggris menggunakan *Support Vector Machine (SVM)* dengan ekstraksi TF-IDF, SVM dengan bantuan N-gram, *Long Short-Term (LSTM)*, *Bidirectional LSTM (Bi-LSTM)*, dan LSTM dengan *GloVe Embedding*. Hasil dari penelitian ini memperlihatkan akurasi terbaik untuk dataset berbahasa India diraih oleh SVM dengan TF-IDF sebesar 87,1% dan Bi-LSTM sebesar 87%, sedangkan untuk dataset berbahasa Inggris diraih oleh Bi-LSTM sebesar 93,4% dan LSTM dengan *GloVe* sebesar 93,3%. Kemudian, terdapat juga penelitian yang berjudul “*Two-Stage Classifier for COVID-19 Misinformation Detection Using BERT: a Study on Indonesian Tweets*” (Faisal & Mahendra, 2022). Penelitian tersebut melakukan dua tahapan klasifikasi secara *pipeline*, yaitu tahapan pertama menyaring *tweet* yang relevan dan selanjutnya tahapan kedua mendeteksi *tweet* misinformasi. Penelitian dilakukan menggunakan *traditional machine learning* (NB, SVM, LR, DT, RF, dan XGB) dengan ekstraksi fitur (BoW, TF-IDF, Word2Vec, dan IndoBERT), serta penelitian juga dilakukan menggunakan *deep learning* (BERT, Bi-LSTM, DNN, dan CNN) dengan IndoBERT *Embedding*. Hasil penelitian menyatakan bahwa akurasi terbaik diraih oleh Bi-LSTM dengan bantuan IndoBERT sebesar 87,02%. Selain itu, terdapat juga penelitian yang berjudul “*Sexism Identification in Social Networks*” (Chaudhary & Kumar, 2023). Penelitian ini melakukan identifikasi pernyataan seksisme dengan dataset berbahasa Inggris dan Spanyol. Penelitian ini menggunakan Bi-LSTM yang menghasilkan nilai F1 sebesar 63,56%.

Berdasarkan latar belakang di atas, maka penulis memutuskan untuk menggunakan algoritma *Bidirectional Long Short-Term* dengan bantuan IndoBERT *Embedding* untuk membangun sebuah sistem berbasis web yang dapat melakukan identifikasi terhadap pernyataan-pernyataan misogini yang tersebar di media sosial. Dengan demikian, penulis memberikan judul penelitian ini berupa “IDENTIFIKASI PERNYATAAN MISOGINI BERDASARKAN KOMENTAR MEDIA SOSIAL MENGGUNAKAN BIDIRECTIONAL LONG SHORT-TERM DAN INDOBERT EMBEDDING”.

Penelitian Terdahulu

No.	Penulis	Judul	Tahun
1.	Punyajoy Saha, Binny Mathew, Pawan Goyal, dan Animesh Mukherjee	<i>Hateminers: Detecting Hate Speech Against Women</i>	2018



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

2.	Maibam Debina Devi dan Navanath Saharia	<i>Misogynous Text Classification Using SVM and LSTM</i>	2020
3.	Arnesa Julia Damanik	Identifikasi Pernyataan Misogini Berbahasa Indonesia Berdasarkan Komentar Youtube Menggunakan <i>GloVe Embedding</i> dan <i>Random Forest Classifier</i>	2021
4.	Rizkyta Shainy Angeline, Dade Nurjanah, dan Hani Nurrahmi	<i>Misogyny Speech Detection Using Long Short-Term Memory and BERT Embeddings</i>	2022
5.	Douglas Raevan Faisal dan Rahmad Mahendra	<i>Two-Stage Classifier for COVID-19 Misinformation Detection Using BERT: a Study on Indonesian Tweets</i>	2022
6.	Atul Chaudhary dan Ritesh Kumar	<i>Sexism Identification in Social Networks</i>	2023

Perbedaan Penelitian

Adapun perbedaan penelitian ini dengan penelitian yang dilakukan oleh Damanik (2021) adalah penelitian tersebut melakukan identifikasi misogini terhadap *file .csv* berisi komentar yang diunggah ke sistem, sedangkan penelitian ini akan melakukan identifikasi misogini melalui komentar-komentar yang diekstraksi dari *link* postingan yang diunggah ke dalam sistem. Selain itu, perbedaan penelitian ini dengan penelitian yang dilakukan oleh Saha et al. (2018), Devi & Saharia (2020), dan Angeline et al. (2022) yang juga membahas identifikasi misogini adalah penelitian ini menggunakan kombinasi *Bidirectional Long Short-Term* sebagai model untuk mengidentifikasi dan *IndoBERT Embedding* sebagai ekstra fitur yang membantu model menganalisis dan mempelajari hubungan kontekstual antar kata dengan lebih baik.

Rumusan Masalah

Kebebasan yang diberikan terhadap pengguna media sosial melalui karakteristik anonimitas dan dalam mengunggah konten telah menjadi salah satu faktor maraknya unggahan yang berisikan hal-hal negatif, termasuk komentar misogini. Dan dalam menyikapi komentar-komentar tersebut, perempuan harus meluangkan banyak waktu untuk dapat memblokir atau melaporkan pengguna yang melakukan pelecehan karena harus memilah mana pernyataan pengguna yang termasuk misogini dan tidak misogini. Hal ini tentunya menyebabkan masyarakat khususnya perempuan merasa tidak aman dan nyaman berada di dunia maya maupun dunia nyata. Serta, dapat memberikan dampak negatif terhadap psikis perempuan yang menjadi korban. Oleh karena itu, diperlukan suatu pemodelan yang secara otomatis dapat mengidentifikasi suatu komentar mengandung pernyataan misogini atau tidak.



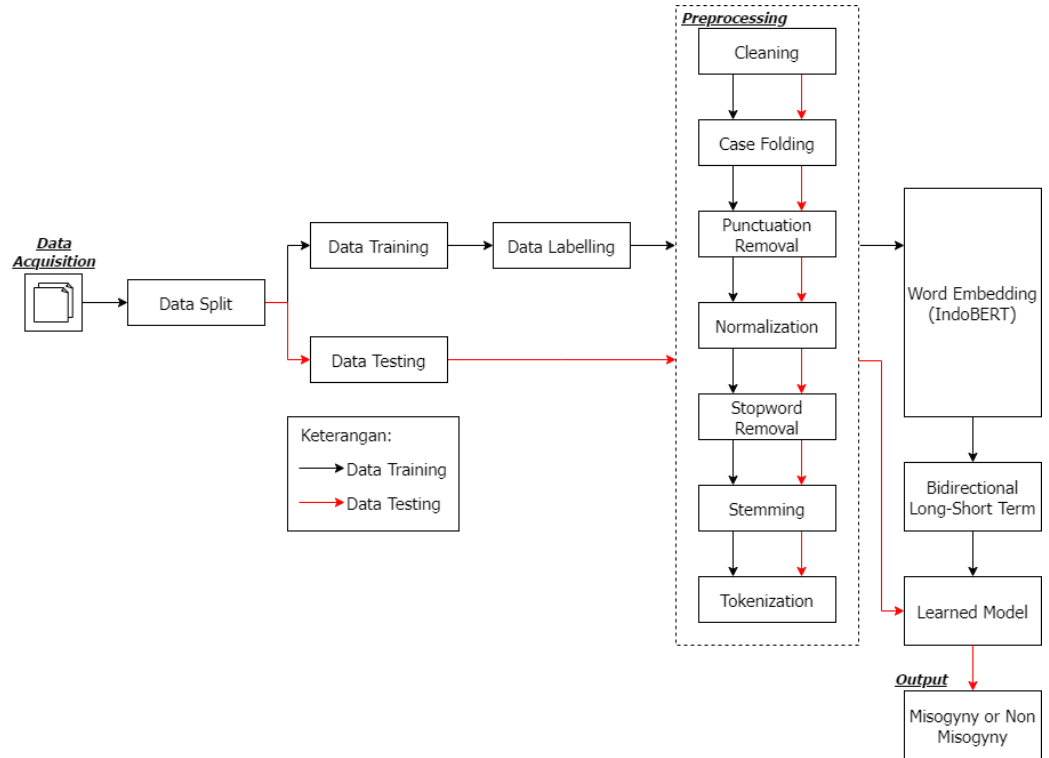
KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: http://it.usu.ac.id

Metodologi



Data Acquisition

Tahapan ini merupakan tahapan awal yang melakukan pengumpulan data-data komentar media sosial. Data komentar diambil dari media sosial Youtube, Instagram, dan X menggunakan teknik *scrapping*. Kemudian, data tersebut dibagi menjadi dua bagian yaitu, data *training* dan data *testing* dengan rasio tertentu.

Labelling

Tahapan ini melakukan proses pemberian label berupa 0 dan 1 untuk setiap data *testing*. Dimana, kelas 0 untuk menyatakan pernyataan non-misogini dan kelas 1 untuk pernyataan misogini.

Preprocessing

Pada tahapan ini, dilakukan serangkaian proses yang bertujuan untuk menghasilkan data yang baik sehingga data dapat lebih mudah dimengerti oleh model. Serangkaian proses tersebut berupa tahapan *cleansing*, *case folding*, *punctuation removal*, *normalization*, *stopword removal*, *stemming*, dan *tokenization*.



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

	<p>Word Embedding</p> <p>Setelah tahap <i>preprocessing data</i> telah selesai, pada tahapan ini akan dilakukan pengubahan data menjadi vektor menggunakan <i>word embedding</i> agar dapat melakukan pemodelan. <i>Word embedding</i> yang digunakan ialah salah satu contoh <i>contextualized word embedding</i> yaitu IndoBERT, dimana vektor dari kata yang sama akan dihasilkan berbeda jika memiliki arti yang berbeda dalam konteks. Representasi vektor yang dilakukan menggunakan IndoBERT ini perlu memberikan inputan sebuah tambahan token [CLS] di awal dan [SEP] di akhir.</p> <p>Modelling</p> <p>Tahapan ini akan melakukan pelatihan model agar dapat mengidentifikasi <i>data train</i> yang telah diubah menjadi vektor menggunakan algoritma <i>Bidirectional Long Short-Term</i>. Algoritma ini bekerja secara dua arah, dimana lapisan pertama (<i>forward layer</i>) akan berfokus untuk urutan setiap kata yang ada pada teks dan lapisan kedua (<i>backward layer</i>) akan merepresentasikan konteks dari kata-kata tersebut. Hasil pelatihan model (<i>learned model</i>) akan digunakan untuk mengidentifikasi komentar.</p> <p>Output</p> <p>Keluaran dari keseluruhan proses tersebut adalah identifikasi komentar yang mengandung unsur misogini dan non-misogini.</p>
Referensi	<p>Angeline, R. S., Nurjanah, D., & Nurrahmi, H. (2022, August). Misogyny Speech Detection Using Long Short-Term Memory and BERT Embeddings. <i>2022 5th International Conference on Information and Communications Technology (ICOIACT)</i>. https://doi.org/https://doi.org/10.1109/ICOIACT55506.2022.9972171</p> <p>Chaudhary, A., & Kumar, R. (2023). Sexism Identification In Social Networks. In M. Aliannejadi, G. Faggioli, N. Ferro, & M. Vlachos (Eds.), <i>CLEF 2023: Conference and Labs of the Evaluation Forum</i> (pp. 891–900). CEUR-WS.org. https://ceur-ws.org/Vol-3497/paper-075.pdf</p> <p>Damanik, A. J. (2021). IDENTIFIKASI PERNYATAAN MISOGINI BERBAHASA INDONESIA BERDASARKAN KOMENTAR YOUTUBE MENGGUNAKAN GLOVE EMBEDDING DAN RANDOM FOREST CLASSIFIER. In <i>Universitas Sumatera Utara</i>.</p> <p>Devi, M. D., & Saharia, N. (2020). Misogynous Text Classification Using SVM and LSTM. In D. Garg, K. Wong, J. Sarangapani, & S. K. Gupta (Eds.), <i>International Advanced Computing Conference 2020</i> (pp. 336–348). Springer Singapore. https://doi.org/https://doi.org/10.1007/978-981-16-0401-0</p> <p>Faisal, D. R., & Mahendra, R. (2022). <i>Two-Stage Classifier for COVID-19 Misinformation Detection Using BERT: a Study on Indonesian Tweets</i>. https://arxiv.org/abs/2206.15359</p> <p>Fuad, B., Amiruddin, M., Yentriyani, A., Kanti, D., Qibtiyah, A., Adkiras, F., Nugroho, H. A., Anshor, M. U., Nahe'i, Novianti, Salampessy, O. C., Hutabarat, R. M., Ratnawati, R., Mashudi, S., Tardi, S. A., Iswarini, T. S. E., Wiandani, T., & Sitohang,</p>



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN
TEKNOLOGI

UNIVERSITAS SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

PROGRAM STUDI S1 TEKNOLOGI INFORMASI

Jalan Alumni No. 3 Gedung C, Kampus USU Padang Bulan, Medan 20155
Telepon/Fax: 061-8210077 | Email: tek.informasi@usu.ac.id | Laman: <http://it.usu.ac.id>

	<p>V. (2023). <i>CATAHU 2023: Kekerasan terhadap Perempuan di Ranah Publik dan Negara: Minimnya Perlindungan dan Pemulihan</i>. https://komnasperempuan.go.id/catatan-tahunan-detail/catahu2023-kekerasan-terhadap-perempuan-di-ranah-publik-dan-negara-minimnya-perlindungan-dan-pemulihan</p> <p>Lubis, F. (2021, October 28). <i>Seksisme dan Misogini dalam Perspektif HAM</i>. Komnas HAM. https://www.komnasham.go.id/index.php/news/2021/10/28/1963/seksisme-dan-misogini-dalam-perspektif-ham.html</p> <p>Maharani, A. P. (2022, November 10). <i>Misogini: Kebencian Ekstrem Hanya Karena Kita Perempuan</i>. Lembaga Pers Mahasiswa Gelora Sriwijaya. https://gelorasriwijaya.co/blog/misogini-kebencian-ekstrem-hanya-karena-kita-perempuan/</p> <p>National Democratic Institute. (2019). <i>Tweets That Chill: Analyzing Online Violence Against Women in Politics</i>. https://www.ndi.org/tweets-that-chill</p> <p>Saha, P., Mathew, B., Goyal, P., & Mukherjee, A. (2018). <i>Hateminers : Detecting Hate Speech Against Women</i>. https://doi.org/https://doi.org/10.48550/arXiv.1812.06700</p>
--	--

Medan, 01 Februari 2024
Mahasiswa yang mengajukan,

(Stephani Uli Basa Silitonga)

NIM. 201402068