

ANALISIS SENTIMEN *BODY SHAMING* MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* PADA MEDIA SOSIAL

Laily Fadlilatul Mi'roj^{1*}, Faisal Reza Pradhana², Triana Harmini³

^{1,2,3}Program Studi Teknik Infomatika, Universitas Darussalam Gontor, Ponorogo, Indonesia

*Penulis korespondensi e-mail : lailymi'roj38@mhs.unida.gontor.ac.id

ABSTRAK

Body shaming adalah tindakan memberikan komentar negatif terhadap fisik seseorang, seperti "gendut", "peseq", atau "cungring", yang sering terjadi di platform media sosial, termasuk Twitter. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap *body shaming* di Twitter menggunakan metode *Support Vector Machine* (SVM). Dataset dikumpulkan melalui teknik *crawling* pada Twitter, kemudian diklasifikasikan menjadi tiga sentimen: positif, netral, dan negatif. Evaluasi model dilakukan menggunakan *confusion matrix*, menghasilkan akurasi sebesar 66%, *precision* 69%, *F1-score* 66%, dan *F1-score* 65%. Distribusi sentimen menunjukkan kelas positif mendominasi dengan 1975 data (40,5%), diikuti kelas netral sebanyak 1829 data (37,5%), dan kelas negatif sebanyak 1071 data (22%). Hasil penelitian menunjukkan bahwa metode SVM cukup efektif untuk mengklasifikasikan sentimen *body shaming* di Twitter. Temuan ini dapat memberikan wawasan dalam pengembangan algoritma deteksi sentimen dan kebijakan moderasi konten di media sosial.

Kata kunci: *Body shaming* , Twitter, *Support Vector Machine* .

1 PENDAHULUAN

Media sosial berkembang pesat tidak hanya di negara-negara maju tetapi juga di negara-negara berkembang seperti Indonesia. Di Indonesia, penggunaannya semakin meluas, bahkan menggantikan peran media tradisional dalam menyebarkan informasi. Indonesia juga memiliki jumlah pengguna Twitter tertinggi kelima di dunia.(Yusuf et al. 2023). Twitter adalah platform media sosial global populer yang memungkinkan pengguna untuk berbagi pemikiran, perasaan, dan pengalaman. Sebagai produk revolusi teknologi, Twitter telah menjadi wadah penting bagi banyak orang untuk menyampaikan pikiran, pendapat, dan kritiknya. Namun, terlepas dari kelebihanannya, platform ini juga memiliki kekurangan, termasuk aspek negatifnya seperti seringnya terjadi *body shaming* di kalangan pengguna.(Sari et al. 2021). *Body shaming* , atau tindakan mengkritik penampilan orang lain tanpa alasan, sudah menjadi praktik umum, terutama di kalangan orang asing. Perilaku ini dianggap sebagai bentuk intimidasi, baik verbal, tertulis, atau psikologis, dan berfokus pada penampilan seseorang. Misalnya mengasosiasikan bagian tubuh tertentu, seperti bentuk hidung, tubuh yang kurus, atau ciri fisik lainnya, dengan stereotip negatif.(Fattah and Purnawansyah 2022)

Perilaku pengguna Twitter berupa opini dan fakta tentang *Body shaming* , jika diolah dengan baik dan benar, dapat menjadi sumber data, menghasilkan informasi yang digunakan untuk meningkatkan persepsi terhadap hal tersebut. Bentuk analisis yang dapat digunakan untuk menganalisis perilaku tersebut adalah analisis sentimen. Analisis sentimen adalah cara yang berguna untuk menganalisis komentar Twitter. Proses ini memungkinkan pemrosesan otomatis dan interpretasi materi teks untuk memperoleh informasi penting. *Support Vector Machine* (SVM)

adalah algoritma yang sering digunakan dalam analisis sentimen dan dikenal efektif mengklasifikasikan data teks.(Husada and Paramita 2021). Salah satu algoritma yang signifikan dan efisien dalam machine learning, termasuk penerapannya dalam analisis sentimen, adalah *Support Vector Machine* (SVM). Algoritma ini berfungsi dengan membuat hyperplane yang mengkategorikan data ke dalam kelompok tertentu, sehingga menghasilkan tingkat akurasi yang berbeda-beda.(Fauzi, Harli, and Kusmanto 2023)

Penelitian tentang analisis sentimen media sosial di Twitter telah dilakukan secara ekstensif di masa lalu dengan menggunakan berbagai teknik analisis yang berbeda. Dalam penelitian yang dilakukan oleh St. Fajriah Fattah, dan Purnawansyah pada tahun 2022 yang berjudul “ Analisis Sentimen terhadap *body shaming* pada twitter menggunakan metode Naïve Bayes Classifier” yang menghasilkan pengujian performansi Accuracy, Precision, *F1-score*, dan *f-measure* dengan featured model trigram menggunakan metode naïve bayes classifier dilakukan pada dataset tweets *body shaming* yang berjumlah 908 data. Berdasarkan hasil pengujian performa dengan model trigram didapatkan hasil Accuracy 61%, *F1-score* 55% dan *f-measure* 55%. Namun Saat ini masih terdapat kekurangan penelitian yang mengkaji sentimen *body shaming* di Twitter Indonesia melalui penggunaan metode *Support Vector Machine* .

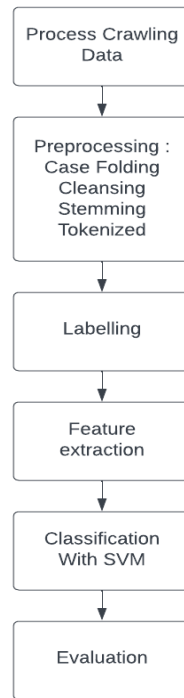
Berdasarkan permasalahan yang telah dijelaskan sebelumnya, penelitian ini akan melakukan analisis sentimen *body shaming* di Twitter menggunakan *Support Vector Machine* . Metode evaluasi yang akan dilakukan adalah dengan menggunakan Confussion Matrix untuk memperoleh akurasi, presisi, *F1-score* dan f-1 score. Penelitian ini membatasi penelitian hanya pada dataset tweet berbahasa Indonesia di media sosial Twitter dengan menggunakan metode klasifikasi *Support Vector Machine* dengan bantuan Twitter auth token.

2 METODE

Metode meliputi prosedur atau tahapan penelitian dan algoritma yang digunakan untuk penelitian. Beberapa hal yang berhubungan dengan tahapan penelitian akan dijelaskan sebagai berikut :

2.1 Tahapan penelitian

Prosedur penelitian menjelaskan berbagai tahapan penelitian yang dilakukan untuk memperoleh hasil akhir analisis sentimen terkait *body shaming* dari tweet pengguna Twitter. Eksperimen penelitian dimulai dengan mengambil data tweet dari Twitter melalui proses yang disebut “*crawling*.” Dengan beberapa penggunaan kata kunci. Data yang telah di *crawling* kemudian diolah menjadi teks preprocessing yang terdiri dari beberapa tahap yaitu case folding, cleansing, stemming, tokenized. Setelah melakukan pemrosesan data langkah selanjutnya adalah proses labelling data dengan menggunakan Library Vader Lexicon data di pisah menjadi 3 kelas yaitu positif, negative , dan netral. Setelah proses pelabelan dilakukan proses pembobotan menggunakan metode TF-IDF dan setelahnya dilakukan klasifikasi menggunakan algoritma *Support Vector Machine* data awal yang diperoleh akan dibagi menjadi data pengujian sebesar 20% dan data pelatihan sebesar 80 %. Lalu selanjutnya mengevaluasi metode klasifikasi dengan prediksi dan *F1-score* yang dibuat oleh *Support Vector Machine* dengan menggunakan *Confusion matrix*. Tahapan penelitian berupa flowchart dapat dilihat pada gambar 1



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Dataset untuk penelitian ini terdiri dari data tweet yang bersumber dari Twitter, yang diperoleh melalui prosedur *crawling* data dengan memanfaatkan twitter auth token. Metode *crawling* data ini melibatkan pengambilan atau pengunduhan informasi dari server Twitter, khususnya berfokus pada data pengguna atau konten tweet berdasarkan kata kunci yang relevan dengan penelitian ini. Untuk mendapatkan data tweet yang diinginkan, sebuah program dikembangkan dengan menggunakan kata kunci tertentu. Contoh *Keyword* yang di cari untuk data perilaku *body shaming* seperti “cungkring”, “gemuk”, “jangkung”, “pendek”, dan lain sebagainya

2.3 Preprocessing

Normalisasi istilah dalam kalimat dikenal sebagai preprocessing. Langkah penting ini memastikan bahwa data pelatihan berkualitas tinggi dan fitur yang diekstraksi selaras dengan hasil yang diinginkan. Pada akhirnya, ini menyederhanakan pemrosesan data secara keseluruhan. Terdapat beberapa langkah dalam tahapan preprocessing yaitu :

2.3.1 Case Folding

Case Folding merupakan tahap dimana karakter alfabet yang telah melalui tahap sanitasi diubah menjadi huruf kecil.

2.3.2 Cleansing

Tahap ini adalah penghapusan karakter non-abjad untuk mengurangi noise. Karakter yang dihilangkan antara lain tanda baca seperti titik (.), koma (,), tanda tanya (?), dan tanda seru (!), serta simbol seperti tanda “@” untuk nama pengguna, hashtag (#), emotikon, dan alamat situs web.

2.3.3 Stemming

Fungsi tahap ini adalah mengubah setiap kata yang dipilih dengan imbuhan menjadi kata dasar

2.3.4 *Tokenizing*

Tahap ini berperan sebagai pemecah kalimat berdasarkan setiap kata (disebut istilah atau token) yang menyusun kalimat tersebut. Tokenisasi dipecah berdasarkan spasi.

2.4 *Labelling*

Pelabelan adalah proses pemberian label atau kategori pada data atau contoh yang digunakan untuk melatih model pembelajaran mesin. Dalam analisis sentimen, pelabelan mengacu pada proses pemberian label atau kategori tertentu pada teks atau komentar berdasarkan sentimen yang terkandung dalam teks atau komentar tersebut. Pelabelan menggunakan VADER (*Valence-Aware Dictionary and Sentiment Reasoner*) adalah algoritma analisis sentimen berbasis leksikon yang dirancang untuk mendeteksi sentimen dalam teks, terutama dalam konteks media sosial, ulasan, atau data teks informal lainnya. Yang mana setiap kata diberi skor sentiment (valence) yang bernilai dari -1 (sangat negatif) hingga +1 (sangat positif). (Hoti and Ajdari 2023) Kesimpulan sentiment didasarkan pada hal berikut:

- a. Sentiment Positive : compound score ≥ 0.05 ;
- b. Sentiment Netral : (compound score > -0.05) = (compound score < 0.05)
- c. Sentiment Negative : compound score ≤ -0.05

2.5 *Feature Extraction*

Sebelum mengklasifikasikan data, setiap kata terlebih dahulu dianalisis secara menyeluruh menggunakan *TfidfVectorizer* dari *sklearn toolkit*. Dengan menggunakan fitur ekstraksi TFIDF dapat meningkatkan efisiensi klasifikasi, yang akan mengurangi waktu pemrosesan komputer dan meningkatkan akurasi.

2.6 *Classification with SVM*

Dalam proses SVM, data awal yang diperoleh akan dibagi menjadi data pengujian dan data pelatihan berdasarkan korelasinya. Pelatihan data mengacu pada data mentah yang kami miliki. Mengenai data pengujian, ini adalah data penilaian yang diperoleh dari data awal tetapi dihasilkan secara acak oleh sistem. Sistem selanjutnya akan menggunakan metode pencarian *grid* untuk menentukan parameter optimal. Selanjutnya sistem akan divalidasi untuk melakukan proses pengujian dan memperoleh hasil yang akurat.

2.7 *Evaluation*

Pada tahapan ini, peneliti akan mengevaluasi metode klasifikasi dengan prediksi dan *F1-score* yang dibuat oleh *Support Vector Machine* dengan menggunakan *Confusion matrix*. Hasil evaluasi akan berguna untuk menentukan apakah model yang digunakan sesuai atau tidak dalam penelitian ini dan penerapannya. *Confusion matrix* merupakan metode yang digunakan untuk mengevaluasi model algoritma klasifikasi yang diterapkan sehingga peserta pelatihan dapat memahami seberapa baik kinerja sistem klasifikasi.

3 HASIL DAN PEMBAHASAN

3.1 *Pengumpulan Data*

Proses pengumpulan data dilakukan dengan menggunakan *Tweet-harvest* untuk *crawling* data dari Twitter. Kata kunci yang digunakan adalah “cebol”, “cungkring”, “jangkung”, “pendek”, “pesek”,

“botak”, “kerempeng”, “tonggos”, “gemuk”. Menghasilkan data mentah yang berjumlah 5006 data dari semua kata kunci yang dicari dari tanggal 1 Mei – 31 Juli 2024.

Tabel 1. Sampel Hasil Pengumpulan Data

no	full_text	created_at	username
1	Walah pesona laki-laki cungring. https://t.co/N5M88KfYr @RizkyDS_ @onthesbrew Padahal udah ga	2024-05-01 00:18:48+00:00	seungminfeb
2	cebol lagi . Itu si ridho aja lebih tinggi dari kiper uzbek. Emang kalah mental dan taktik aja yes. Lawan iraq ntar kalah nasab	2024-05-01 00:30:33+00:00	lupa_mancing
3	kok gua jd cebol https://t.co/NYm5btSYpQ	2024-05-01 01:23:52+00:00	shxxbiaii

3.2 Labelling

Setelah data berhasil dikumpulkan, langkah selanjutnya adalah memastikan kualitas data dengan menghapus data duplikat. Analisis sentimen dilakukan dengan menggunakan NLTK Library Vader Lexicon. Library ini bekerja dengan menganalisis setiap kata dalam data, memberikan skor pada setiap kata, lalu menghitung skornya. Hasil akhir digunakan untuk menentukan apakah data tersebut negatif, positif, atau netral. Setelah pelabelan dengan kamus lexicon dilakukan validasi untuk menghindari kesalahan dalam pelabelan. Hasil dari pada pelabelan data twitter dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pelabelan Data Twitter

no	sentence	scores	compound	sentiment
1	walah pesona laki laki cungring	{'neg': 0.0, 'neu': 0.381, 'pos': 0.619, 'compound': 0.7579}	0.7579	Positive
2	cebol ridho kiper uzbek kalah mental taktik yes lawan iraq entar kalah nasab	{'neg': 0.282, 'neu': 0.552, 'pos': 0.166, 'compound': -0.2263}	-0.2263	Negative
3	gua cebol	{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}	0	Neutral

Rincian jumlah data untuk setiap kelas disajikan pada Tabel 7. Dari data tersebut terlihat bahwa kelas positif mendominasi dengan total 1.975 data (40,5%), diikuti oleh kelas netral sebanyak 1.829 data (37,5%), dan kelas negatif yang merupakan jumlah terkecil, yaitu 1.071 data (22%). Hal ini mencerminkan distribusi sentimen yang lebih condong ke arah positif.

3.3 Preprocessing

a. Case folding

Case Folding merupakan tahap dimana karakter alfabet yang telah melalui tahap sanitasi diubah menjadi huruf kecil. Sebagai contoh dari case folding adalah “Perempuan” setelah di proses menjadi “perempuan”

Tabel 3. Sampel Hasil Case Folding

no	data awal	case_folding
1	Walah pesona laki-laki cungkring. https://t.co/N5M88KFkYr Padahal udah ga cebol lagi . Itu si ridho aja lebih tinggi	walah pesona laki laki cungkring padahal udah cebol lagi itu ridho aja lebih tinggi
2	dari kiper uzbek. Emang kalah mental dan taktik aja yes. Lawan iraq ntar kalah nasab	dari kiper uzbek emang kalah mental dan taktik aja yes lawan iraq ntar kalah nasab
3	kok gua jd cebol https://t.co/NYm5btSYpQ	kok gua cebol

b. Cleansing

Tahap ini adalah penghapusan karakter non-abjad untuk mengurangi noise. Karakter yang dihilangkan antara lain tanda baca seperti titik (.), koma (,), tanda tanya (?), dan tanda seru (!), serta simbol seperti tanda “@” untuk nama pengguna, hashtag (#), emotikon, dan alamat situs web.

Tabel 4. Sampel Hasil Cleansing

no	data awal	cleansing
1	Walah pesona laki-laki cungkring. https://t.co/N5M88KFkYr @RizkyDS_ @onthesbrew Padahal udah ga cebol lagi . Itu si ridho aja lebih tinggi	Walah pesona laki laki cungkring Padahal udah cebol lagi Itu ridho aja lebih tinggi
2	dari kiper uzbek. Emang kalah mental dan taktik aja yes. Lawan iraq ntar kalah nasab	dari kiper uzbek Emang kalah mental dan taktik aja yes Lawan iraq ntar kalah nasab
3	kok gua jd cebol https://t.co/NYm5btSYpQ	kok gua cebol

c. Stemming

Fungsi tahap ini adalah mengubah setiap kata yang dipilih dengan imbuhan menjadi kata dasar

Tabel 5. Sampel Hasil Stemming

no	cleansing	stemming
1	Walah pesona laki laki cungkring	['walah', 'pesona', 'laki', 'laki', 'cungkring']
2	Padahal udah cebol lagi Itu ridho aja lebih tinggi dari kiper uzbek Emang kalah mental dan taktik aja yes Lawan iraq ntar kalah nasab	['cebol', 'ridho', 'kiper', 'uzbek', 'kalah', 'mental', 'taktik', 'yes', 'lawan', 'iraq', 'entar', 'kalah', 'nasab']
3	kok gua cebol	['gua', 'cebol']

d. Tokenizing

Tahap ini berperan sebagai pemecah kalimat berdasarkan setiap kata (disebut istilah atau token) yang menyusun kalimat tersebut. Tokenisasi dipecah berdasarkan spasi.

Tabel 6. Sampel Hasil Tokenizing

no	stemming	tokenisasi
1	['walah', 'pesona', 'laki', 'laki', 'cungkring']	['walah', 'pesona', 'laki', 'laki', 'cungkring']
2	['cebol', 'ridho', 'kiper', 'uzbek', 'kalah', 'mental', 'taktik', 'yes', 'lawan', 'iraq', 'entar', 'kalah', 'nasab']	['padahal', 'udah', 'cebol', 'lagi', 'itu', 'ridho', 'aja', 'lebih', 'tinggi', 'dari', 'kiper', 'uzbek', 'emang', 'kalah', 'mental', 'dan', 'taktik', 'aja', 'yes', 'lawan', 'iraq', 'ntar', 'kalah', 'nasab']
3	['gua', 'cebol']	['kok', 'gua', 'cebol']

3.4 Feature Extraction

Data yang telah melalui proses preprocessing diberikan bobot pada tahap pembobotan. Teknik pembobotan yang digunakan dalam penelitian ini adalah TF IDF, yang merupakan salah satu teknik yang umum digunakan dalam analisis data sentimen. Pembobotan dilakukan untuk mengatasi ketidakseimbangan kata dengan memberikan bobot yang sesuai untuk kata-kata yang memiliki dampak kuat saat dilakukan analisis sentimen. Hasil pembobotan TF-IDF di tunjukkan pada Tabel 7

Tabel 7. Hasil Pembobotan TF-IDF

Dokumen	Indeks	Nilai TF-IDF
0	4893	0.571149533737192
1	9732	0.185947383771401
2	5486	0.506053202629998

3.5 Classification With SVM

Setelah menyelesaikan tahap preprocessing, data dipecah menjadi dua bagian utama, yaitu data pelatihan dan data uji, dengan perbandingan 80:20. Sebanyak 3.901 data digunakan untuk melatih model, sedangkan 975 data sisanya digunakan untuk menguji performa model. Dalam penelitian

ini, Kernel LinearSVC diterapkan sebagai metode utama pada algoritma *Support Vector Machine* (SVM) untuk menganalisis data secara lebih akurat.

3.6 Pengujian Model SVM dengan *Confusion matrix*

Model *Support Vector Machine* (SVM) yang dikembangkan kemudian dievaluasi menggunakan metode *Confusion matrix*. Tabel 8 memperlihatkan bahwa nilai *precision* kelas negative sebesar 0.60, *F1-score* sebesar 0.38, dan *F1-score* sebesar 0.47, lalu nilai *precision* kelas netral sebesar 0.59, *F1-score* 0.92, dan *F1-score* sebesar 0.72, kemudian nilai *precision* kelas positive sebesar 0.83, *F1-score* 0.57 dan *F1-score* sebesar 0.68. nilai akurasi keseluruhan yang didapat adalah 0.66.

Tabel 8. Hasil Akurasi SVM

	<i>precision</i>	<i>F1-score</i>	f1 score	support
Negatif	0.60	0.38	0.47	211
Netral	0.59	0.92	0.72	362
Positif	0.83	0.57	0.68	402
Macro Avg	0.67	0.63	0.62	975
Weighted Avg	0.69	0.66	0.65	975
Accuracy			0.66	975

Hasil penelitian menggunakan algoritma *Support Vector Machine* ini sedikit lebih tinggi dibandingkan dengan penelitian sebelumnya yang menggunakan algoritma naïve bayes yaitu dengan akurasi 61% (Fattah and Purnawansyah 2022)

4 KESIMPULAN

Dari hasil penelitian diatas dapat disimpulkan bahwa

- Menghasilkan data tweet *body shaming* sebesar 1.975 (40.5%) pada kelas sentiment positif, 1829 (37.5%) pada kelas netral, dan yang terakhir 1071 (22%) pada kelas negatif
- Sentiment analisis *body shaming* menggunakan algoritma *Support Vector Machine* menghasilkan nilai akurasi sebesar 66% dengan nilai *precision* sebesar 69%, *F1-score* sebesar 66% dan *F1-score* sebesar 65%
- Algoritma *Support Vector Machine* mendapatkan nilai akurasi lebih tinggi dalam analisis sentiment *body shaming* dibanding dengan penelitian sebelumnya yang menggunakan algoritma naïve bayes.

UCAPAN TERIMA KASIH

Ucapan terima kasih penulis sampaikan kepada Dosen Pembimbing yang telah memberikan arahan, bimbingan, dan masukan yang sangat berharga selama proses penelitian ini. Tidak lupa, penulis mengucapkan terima kasih kepada Validator Data, yang telah berkontribusi dalam memverifikasi dan memastikan kualitas data, serta kepada semua pihak yang telah memberikan dukungan baik secara langsung maupun tidak langsung dalam menyelesaikan penelitian ini.

DAFTAR PUSTAKA

Fattah, St. Fajriah, and Purnawansyah. 2022. "Analisis Sentimen Terhadap *Body shaming* Pada Twitter Menggunakan Metode Naïve Bayes Classifier." *Indonesian Journal of Data and Science* 3 (2): 61–71. <https://doi.org/10.56705/ijodas.v3i2.46>.

- Fauzi, Ahmad, Eko Harli, and Tria Hadi Kusmanto. 2023. "Analisis Sentimen Wacana Penerapan Jalan Berbayar Di Jakarta Menggunakan Algoritma SVM." *Resolusi: Rekayasa Teknik Informatika* ... 3 (6): 430–35. <http://djournals.com/resolusi/article/view/1054%0Ahttp://djournals.com/resolusi/article/download/1054/622>.
- Hoti, Mërgim H., and Jaumin Ajdari. 2023. "Sentiment Analysis Using the Vader Model for Assessing Company Services Based on Posts on Social Media." *SEEU Review* 18 (2): 19–33. <https://doi.org/10.2478/seeur-2023-0043>.
- Husada, Hendry Cipta, and Adi Suryaputra Paramita. 2021. "Analisis Sentimen Pada Maskapai Penerbangan Di Platform Twitter Menggunakan Algoritma *Support Vector Machine* (SVM)." *Teknika* 10 (1): 18–26. <https://doi.org/10.34148/teknika.v10i1.311>.
- Sari, Septika Sari, Ulfa Khaira Ulfa, Pradita Eko Prasetyo Utomo Pradita, and Tri Suratno Tri. 2021. "Analisis Sentimen Terhadap Komentar Beauty Shaming Di Media Sosial Twitter Menggunakan Algoritma SentiStrength." *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)* 1 (1): 71–78. <https://doi.org/10.57152/ijirse.v1i1.55>.
- Yusuf, Faidah, Hardianto Rahman, Sitti Rahmi, and Angri Lismayani. 2023. "Pemanfaatan Media Sosial Sebagai Sarana Komunikasi, Informasi, Dan Dokumentasi: Pendidikan Di Majelis Taklim Annur Sejahtera." *JHP2M: Jurnal Hasil-Hasil Pengabdian Dan Pemberdayaan Masyarakat* 2: 1–8.