PROSIDING SEMINAR NASIONAL PEMANFAATAN SAINS DAN TEKNOLOGI INFORMASI 2023
Vol. 1 No. 1 (2023)97-102

Sentimen Analisis Pertambangan Di Konawe Utara Dengan Metode Naïve Bayes

Budi Wijaya Rauf
[1]Universitas Muhammadiyah Kendari, Kendari
e-mail: **budi.wijaya@umkendari.ac.id**

Abstrak

Sulawesi Tenggara merupakan salah satu provinsi yang ada di Indonesia yang terkenal akan daerah pertambangannya. Salah satu kabupaten/kota yang terkenal akan pertambangannya adalah Konawe Utara. Blok Mandiodo yang terletak di desa Mandiodo kecamatan Molawe kini menarik perhatian masyarakat Indonesia karena kasus korupsi. Banyak pendapat dan opini pun beredar terkait pertambangan yang dijadikan lahan korupsi ini, opini masyarakat dapat bersifat netral, positif, maupun negatif. Salah satu wadah untuk menuangkan pikiran dan opini terhadap kasus tersebut adalah sosial media, Twitter. Penelitian ini dilakukan untuk menganalisis sentimen masyarakat terhadap pertambangan yang ada di Konawe Utara, sebagaimana karena kasus ini mencuat akan membuat pandangan masyarakat berubah akan pertambangan ini. Tahapan proses untuk melakukan penelitian ini meliputi tahap pra-proses yaitu tahap pengambilan data, tahap pembersihan data yang juga memiliki beberapa langkah yaitu *stopword*, *filtering*, *stemming* dan tokenisasi. Selanjutnya adalah perhitungan kata yang sering muncul dengan TF-IDF lalu merancang model *machine learning* dengan algoritma Naïve Bayes. Hasilnya mendapatkan nilai akurasi sebesar 85%, nilai presisi sebesar 83,5% dan *recall* 87,5%.

Kata kunci— Konawe Utara, *Machine Learning,* Naïve Bayes, Pertambangan, Sentimen Analisis

Abstract

Sulawesi Tenggara is one of the provinces in Indonesia that is known for its mining areas. One of the districts/cities famous for its mining is North Konawe. The Mandiodo Block located in Mandiodo village, Molawe sub-district, has recently captured the attention of the Indonesian public due to a corruption case. Various opinions and perspectives circulate regarding the mining activities that have become a subject of corruption. Public opinions can be neutral, positive, or negative. One of the platforms for expressing thoughts and opinions about this case is Twitter. This research is conducted to analyze the sentiments of the community regarding the mining activities in North Konawe. As this case has gained attention, it has the potential to shift public perspectives on mining. The research process involves several stages, including the pre-processing phase, which includes data collection and data cleansing steps such as stopword removal, filtering, stemming, and tokenization. Subsequently, the frequently occurring words are calculated using TF-IDF (Term Frequency-Inverse Document Frequency). A machine learning model is then designed using the Naïve Bayes algorithm. The results show an accuracy score of 85%, precision of 83.5%, and recall of 87.5%.

Keywords— *Analysis Sentiment, North Konawe, Machine Learning, Naïve Bayes, Mining*

1. PENDAHULUAN

Pertambangan merupakan kegiatan pengambilan endapan bahan galian yang memiliki nilai ekonomis di dalam permukaan bumi, pengambilan tersebut dapat dilakukan secara mekanis maupun secara manual. Hasil dari pertambangan dapat berupa logam mulia, batu mineral maupun minyak. Proses pertambangan ini disisi lain juga dapat merusak alam sehingga di beberapa tempat pertambangan yang baru dibuka terjadi protes dengan membawa isu lingkungan.

Dinamika akan pertambagan membawa hal positif dan negatif terhadap masyarakat yang tinggal di dekat kawasan tersebut. Hal positif dari adanya pembukaan sektor pertambangan di suatu daerah dapat menaikkan kesejahteraan dari masyarakat lokal, mulai dari naiknya pendapatan daerah hingga terserapnya warga lokal sebagai pekerja pada sektor tersebut. Dampaknya adalah perputaran ekonomi di daerah menjadi naik sehingga infrastruktur daerah menjadi maju serta angka kemiskinan pun menurun. Tetapi disisi lain, dampak buruk yang dihasilkan atas pertambangan ini dapat merusak lingkungan yang akan berakibat fatal untuk generasi selanjutnya.

Analisis sentimen telah menjadi alat yang efektif dalam menggali pandangan, opini, dan perasaan masyarakat terkait suatu topik tertentu. Dalam konteks pertambangan, analisis sentimen dapat membantu dalam mengidentifikasi bagaimana masyarakat merespons berbagai aspek pertambangan, termasuk dampak lingkungan, manfaat ekonomi, keberlanjutan, dan isu-isu terkait lainnya. Dengan memahami sentimen yang ada, pihak berkepentingan, termasuk pemerintah, perusahaan pertambangan, serta masyarakat umum, dapat lebih efektif berkomunikasi dan berkolaborasi dalam rangka m encapai tujuan yang saling menguntungkan.

Dalam kaitannya dengan wilayah Konawe Utara, yang dikenal dengan potensi sumber daya tambangnya, analisis sentimen menjadi semakin relevan. Pada satu sisi, potensi pertambangan dapat menjadi peluang untuk pembangunan ekonomi dan peningkatan kesejahteraan masyarakat. Namun, di sisi lain, kegiatan pertambangan juga dapat menimbulkan keprihatinan terhadap dampak lingkungan dan sosial yang mungkin terjadi.

Beberapa penelitian sebelumnya telah mengaplikasikan analisis sentimen dalam konteks pertambangan dan pengelolaan sumber daya alam. Misalnya, penelitian oleh Smith et al. (2018) menginvestigasi sentimen masyarakat terhadap industri pertambangan di wilayah Australia menggunakan pendekatan analisis teks dan algoritma klasifikasi. Hasil penelitian tersebut memberikan wawasan tentang dinamika antara dampak ekonomi dan lingkungan dari kegiatan pertambangan[1]. Di sisi lain, penelitian oleh Johnson et al. (2020) menerapkan metode analisis sentimen untuk mengukur opini publik terkait pertambangan dan dampaknya terhadap komunitas lokal di negara-negara berkembang[2].

Penelitian serupa turut dilakukan oleh Putri et al. (2021) menganalisis sentimen masyarakat terhadap kinerja DPR dengan menggunakan Naïve Bayes dan memiliki akurasi sebesar 80%[3]. Penelitan relevan lainnya dipublikasikan oleh Hasri & Alita (2022) dengan menggunakan dua metode yaitu Naïve Bayes dan *Support Vector Machine*, hasilnya Naïve Bayes memiliki nilai akurasi yang lebih baik yaitu sebesar 81,07% sedangkan SVM sebesar 79%[4].

Sentimen analisis juga dilakukan oleh Buntoro (2017) dengan menggunakan metode Naïve Bayes dalam melakukan klasifikasi sentiment pada calon gubernur DKI Jakarta 2017. Penelitian ini membandingkan antara Naïve Bayes dan *Support Vector Machine* dengan variabel sentimen yaitu AHY, Ahok, dan Anies[5]. Selanjutnya, penelitian terhadap sentimen masyarakat dengan menggunakan media Twitter sebagai tempat pengambilan data juga dilakukan oleh Styawati(2021) menggunakan metode Support Vector Machine dalam penelitiannya[6].

Penelitian relevan lainnya dilakukan oleh Giovani et al (2020) yang menerapkan metode klasifikasi untuk meneliti sentimen masyarakat di media sosial Twitter terhadap aplikasi Ruang Guru. Penelitian ini menerapkan 4 metode klasifikasi yaitu Naïve Bayes, SVM, KNN dan PSO[7].

Penelitian terkait tentang pertambangan yang ada di Konawe Utara juga pernah dilakukan oleh Munawir et al (2023) yang membahas tentang problematika pelaksanaan tanggung jawab sosial perusahaan tambang yang beroperasi di desa Mandiodo, Konawe Utara. Hasil dari penelitian ini menunjukkan

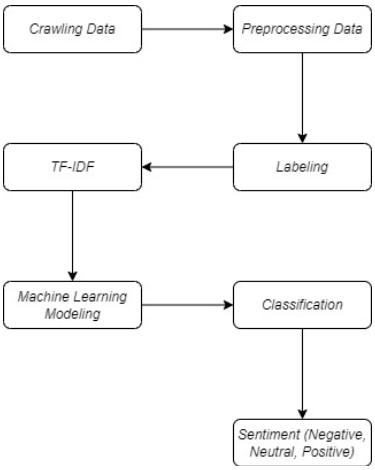
bahwa terjadi pengabaian tanggung jawab terhadap tanggung jawab sosial yang seharusnya dijalankan oleh perusahaan tambang sekitar[8].

Masalah lain terkait pertambangan di Konawe Utara juga pernah diteliti oleh Munawir et al (2023) yang membahas tentang pelanggaran hukum yang dilakukan oleh dua perusahaan tambang yang berada di blok Mandiodo[9].

Dengan begitu, mengacu kepada penelitian relevan yang telah dijabarkan diatas, maka penelitian sentimen analisis terhadap pertambangan di daerah konawe utara perlu dilakukan karena banyak masalah yang terjadi pada pertambangan di Konawe Utara. Dalam penelitian ini akan menggunakan metode analisis sentimen yang menggabungkan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan algoritma klasifikasi Naive Bayes. Metode TF-IDF akan digunakan untuk menganalisis teks-teks yang berkaitan dengan pertambangan di Konawe Utara, sementara algoritma Naive Bayes akan membantu mengklasifikasikan sentimen-sentimen yang terkandung dalam teks-teks tersebut menjadi positif, negatif, atau netral. Kombinasi kedua teknik ini diharapkan dapat memberikan wawasan yang lebih mendalam tentang bagaimana masyarakat merespons pertambangan di wilayah tersebut.

2. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan analisis sentimen menggunakan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) dan algoritma klasifikasi Naive Bayes untuk menganalisis sentimen masyarakat terkait pertambangan di Konawe Utara melalui data cuitan di Twitter.



Gambar 1 *Flowchart* penelitian

Pada gambar 1 dapat dilihat *flowchart* untuk penelitian ini. Proses penelitian berlangsung dengan 7 tahapan dimana tahapannya adalah *crawling data, preprocessing data*, *labeling, Ekstraksi TF-IDF, Machine Learning Modeling, Classification* dan terakhir adalah hasil klasifikasi yaitu sentimen terhadap cuitan.

2.1 *Crawling Data*

Data cuitan terkait tentang pertambangan di Konawe Utara diambil dari Twitter menggunakan *library tweet-harvest* yang dapat mengambil data cuitan menggunakan *auth_token* dan dengan kata kunci yang relevan seperti "Pertambangan Konawe Utara". Pengambilan data dilakukan dalam rentang waktu 2012 hingga 2023 untuk memastikan representasi yang presisi dan akurat.

Data cuitan yang berhasil diperoleh dengan metode ini adalah sebanyak 229 cuitan dan 5 diantaranya dapat dilihat pada tabel 1 berikut ini.

Tabel 1 Contoh Cuitan Hasil *Crawling Data* di Twitter

No.	Cuitan	Tanggal
1	Kejaksaan Tinggi Sulawesi Tenggara menyita uang tunai senilai Rp79 miliar yang merupakan barang bukti korupsi kejahatan pertambangan di wilayah Konawe Utara. https://t.co/fSdh2dCADm	25 Agustus 2023
2	Fantastis, Duit Hasil Korupsi Tambang Nikel di Konawe yang Disita Kejati Rp79 MIliar: Korupsi pertambangan di Blok Mandiodo Konawe Utara terus ditindaklanjuti Kejaksaan Tinggi Sulawesiâ€¦ #Ekonomi #tambangkonawe #BlokMandiodoKonaweUtara #korupsitambang https://t.co/WDp6p2fLP8	25 Agustus 2023

No.	Cuitan	Tanggal
3	Korupsi Pertambangan di Blok Mandiodo Konawe Utara, Kejati Sultra Sita Rp79 Miliar https://t.co/R75amNA16t	24 Agustus 2023
4	Tumpukan Uang Rp 79 Miliar Hasil Sitaan Korupsi Pertambangan di Konawe Utara #TempoVideo https://t.co/ERmkDf5Z2i	24 Agustus 2023
5	Kejaksaan Tinggi (Kejati) Sulawesi Tenggara (Sultra) menyita uang senilai Rp79 miliar dari kasus tindak pidana korupsi pertambangan nikel di wilayah IUP PT Antam Blok Mandiodo, Konawe Utara. https://t.co/Y7OYxqKtqt	24 Agustus 2023

2.2 *Preprocessing Data*

Pada tahapan *preprocessing data* (pra-pemrosesan data) akan dilakukan dengan 4 langkah yaitu *cleansing data, stemming data, stopwords dan tokenisasi*.

Langkah pertama yaitu *cleansing data* atau pembersihan data. Sebagaimana contoh cuitan pada tabel 1, dimana kolom cuitan berisi opini Masyarakat terhadap pertambangan di Konawe Utara. Cuitan-cuitan tersebut memiliki banyak karakter yang tidak diperlukan dalam proses selanjutnya maka harus dibersihkan. Dengan menggunakan *library re*, langkah ini dapat membersihkan karakter *mentions* '@', menghapus *hash tag* '#', menghapus *hyperlink* 'https', dan menghapus tanda baca.

Langkah kedua yaitu proses *stemming data*, langkah ini akan mengembalikkan sebuah kata kembali ke kata dasarnya. *Stemming Data* dapat dilakukan dengan menggunakan *library* Sastrawi yang dapat mengembalikan kata Bahasa Indonesia ke bentuk asalnya.

Tahap selanjutnya adalah *stopwords*. Langkah ini akan menghilangkan seluruh kata-kata umum yang sering digunakan, tetapi tidak memiliki makna jika berdiri sendiri. *Stopwords* merupakan serangkaian daftar kata-kata yang sedikit atau tidak memiliki makna dan jika tidak tidak dihilangkan maka akan mengganggu hasil analisis sentimen. Daftar *stopwords* berasal dari *library nltk* dan merupakan *stopwords* berbahasa Indonesia.

Lalu yang terakhir adalah tokenisasi. Tahap ini akan memecah kalimat cuitan menjadi kata perkata. Hal ini dilakukan untuk memudahkan analisis dalam melakukan pelabelan sentimen. Langkah tokenisasi menggunakan *library nltk* dan menggunakan fungsi *WordPunctTokenizer*() yang memiliki kegunaan untuk memisahkan setiap kata dengan pemisah spasi serta akan menghilangkan tanda baca yang melekat pada sebuah kata yang diantaranya masih lolos pada tahap pembersihan data.

2. 3 TF-IDF

Setelah pra-proses data maka langkah selanjutnya adalah ekstraksi fitur menggunakan metode TF-IDF. Teknik ini adalah pendekatan yang umum digunakan dalam analisis teks untuk menghitung pentingnya kata-kata dalam suatu dokumen dalam kontek koleksi dokumen yang lebih besar. Data cuitan yang berjumlah 229 data akan dilakukan langkah-langkah ekstraksi, diantara lain seperti *Term Frequency* (TF) yaitu sebuah langkah dimana akan menghitung frekuensi kata muncul dari sebuah cuitan.

Selanjutnya adalah *Inverse Document Frequency* (IDF) yaitu sebuah metrik perhitungan yang mengukur seberapa jarang suatu kata muncul dalam sebuah koleksi cuitan. Lalu langkah ketiga adalah menghitung bobot kata yang ada dalam masing-masing cuitan.

$$tf\ idf\ (t, d, D) = tf(t, d).idf(t.D) \quad (1)$$

2. 4 *Labeling*

Tahapan selanjutnya merupakan *labeling* yang memberikan label positif, negatif maupun netral terhadap dataset Twitter. *Labeling* dilakukan dengan mencocokan kalimat dengan kata-kata positif maupun negatif berdasarkan aturan dari KBBI.

2. 4 Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi yang masuk dalam kategori *supervised learning*. Algoritma ini digunakan untuk membuat model *machine learning* yang lalu akan mengklasifikasikan data cuitan.

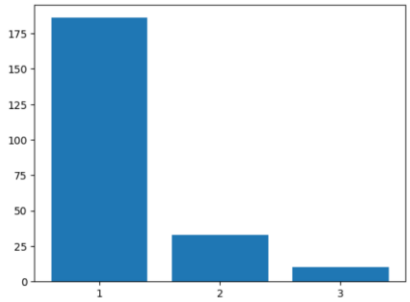
Dataset yang digunakan adalah dataset yang telah dilabeli pada proses sebelumnya.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)} \qquad (2)$$

Nilai dari pembobotan TF-IDF dan juga label dari proses *labeling* akan digunakan sebagai variabel dalam mengklasifikasikan sentimen negatif, netral maupun positif dari sebuah data cuitan. Masing-masing dari setiap data akan dihitung probabilitasnya dan probabilitas tertinggi untuk tiap sentimen akan digunakan sebagai indikator klasifikasi. *Library* yang digunakan adalah *GaussianNB* yang merupakan *library* yang dikembangkan oleh Scikit Learn.

3. HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan sebuah model *machine learning* yang dapat mengklasifikasikan sentimen positif, netral maupun negatif dengan kata kunci relevan seperti ‘pertambangan’ atau ‘Konawe Utara’. Dari 229 data cuitan, terdapat 186 data dengan sentimen negatif, 33 data dengan sentimen netral, dan 10 data dengan sentimen positif, data tersebut dapat dilihat pada gambar 2 berikut.



Gambar 2 Perbandingan sentimen

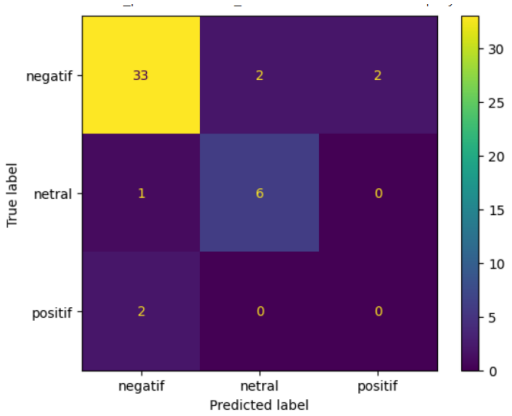
Dari gambar 2 diatas, untuk label 1 merepresentasikan sentimen negatif, label 2 untuk sentimen netral, dan label 3 untuk sentimen positif. Berdasarkan data tersebut dapat terlihat perbandingan antara sentimen negatif, netral dan positif, dimana jumlah sentimen negatif terhadap total data berjumlah 84%. Data ini lalu dibagi menjadi dua bagian, yaitu data latih dan data uji dimana data latih berjumlah 80% dan data uji 20%.

Hasil dari klasifikasi lalu dimasukkan ke dalam *confusion matrix* untuk menghitung tingkat akurasi, presisi dan recall dari algoritma *naïve bayes* dalam menganalisis sentimen masyarakat di Twitter mengenai pertambangan di Konawe Utara. *Confusion matrix* dapat dilihat pada tabel 2 berikut ini.

Tabel 2 *Confusion Matrix*

	Prediksi		
Aktual	Negatif	Netral	Positif
Negatif	33	2	2
Netral	1	6	0
Positif	2	0	0

Hasilnya juga dapat dilihat pada gambar 3 berikut.



Gambar 3 *Confusion Matrix*

Dari hasil perhitungan *confusion matrix*, algoritma Naïve Bayes dapat menghasilkan tingkat akurasi sebesar 85%, presisi 83,5% dan *recall 87,5%*. Ini mengindikasikan bahwa model ini secara keseluruhan memiliki kemampuan yang baik dalam mengklasifikasikan sentimen-sentimen berdasarkan kata kunci yang relevan. Selayaknya seperti referensi penelitian relevan terdahulu yang mana memiliki tingkat akurasi yang serupa. Tetapi, pada penelitian ini dapat terlihat bahwa Naïve Bayes memiliki tingkat akurasi yang memadai bahkan ketika data latih yang digunakan memiliki jumlah yang sedikit dan bias terhadap sentimen negatif. Bagaimanapun, analisis lebih mendalam dapat menggali aspek-aspek seperti *false positive* dan *false negative*, serta interpretasi dari tingkat presisi dan *recall* yang mencerminkan efektivitas model dalam mengatasi masing-masing jenis sentimen.

Hasil penelitian ini memberikan pandangan yang lebih dalam tentang

bagaimana model machine learning dapat diterapkan untuk menganalisis sentimen masyarakat di media sosial terkait topik tertentu, seperti pertambangan di Konawe Utara. Selain itu, hasil evaluasi ini juga dapat menjadi dasar untuk pengembangan model yang lebih canggih dan lebih akurat dalam mengklasifikasikan sentimen-sentimen yang lebih kompleks.

4. KESIMPULAN

Pertambangan di Konawe Utara mendapatkan sentimen negatif sebanyak 186 data dari total 229. Hal ini menunjukan bahwa setidaknya 84% cuitan yang ada di Twitter memiliki sentimen negatif terhadap pertambangan yang ada di Konawe Utara. Selain itu, Algoritma Naïve Bayes juga berhasil mengklasifikasikan sentimen dengan tingkat akurasi 85% dan presisi sebesar 83,5%. Penilitian ini menunjukan bahwa Naïve Bayes mampu memiliki tingkat akurasi yang tinggi meskipun data yang digunakan merupakan data bias terhadap sentimen negatif.

5. SARAN

Penelitian ini dapat dikembangkan ke tahap implementasi ke dalam sebuah website yang dapat secara langsung mengklasifikasikan sentimen dari sebuah cuitan. Selain itu, data yang digunakan sebagai data latih masih sangat sedikit sehingga akan lebih baik jika data cuitan yang digunakan untuk data latih ditambah jumlahnya.

UCAPAN TERIMA KASIH

Saya ucapkan terima kasih kepada rekan dosen Sistem dan Teknologi Informasi (STI) yang membantu dalam mencari data, ucapan terima kasih juga disampaikan kepada penyelenggara Universitas Sembilanbelas November Kolaka yang merangkaikan kegiatan SEMPATIN Call for Paper.

DAFTAR PUSTAKA

[1] K. Smith, A. Kerr, and L. Hopkins, “Public Perceptions of Mining in Australia: What Do Polls Tell Us?,” *Resources*, vol. 7, no. 3, p. 49, 2018.

[2] G. Johnson, T. Cooke, and A. L. Yang, “Public Sentiment towards Mining and Its Implications for Community Engagement,” *Resources*, vol. 9, no. 12, p. 139, 2020.

[3] D. Duei Putri, G. F. Nama, and W. E. Sulistiono, “Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan. 2022, doi: 10.23960/jitet.v10i1.2262.

[4] C. F. Hasri and D. Alita, “Penerapan Metode Naïve Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter,” *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 3, no. 2, pp. 145–160, 2022, [Online]. Available: http://jim.teknokrat.ac.id/index.php/informatika

[5] G. A. Buntoro, “Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter,” 2017. [Online]. Available: https://t.co/jrvaMsgBdH

[6] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” vol. 6, no. 3, 2021, [Online]. Available: http://situs.com

[7] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” *Jurnal Teknoinfo*, vol. 14, no. 2, p. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679.

[8] Munawir, L. O. ., Tolo, S. B. ., & Muin, F. “Problematika Pelaksanaan Tanggungjawab Sosial Perusahan Tambang Terhadap Masyarakat Sekitar Pertambangan Di Desa Mandiodo Kabupaten Konawe Utara: Masyarakat sekitar Pertambangan di Desa Mandiodo Kabupaten Konawe Utara”. *Jurnal Risalah Kenotariatan*, *4*(1). https://doi.org/10.29303/risalahkenotariatan.v4i1.121

[9] Munawir, L. O., Bariun, L., O., Tolo, S. B., & Amir, M. K, “Penegakan Hukum Administrasi Pengelolaan Lingkungan Hidup di Lokasi Pertambangan,” *Widya Yuridika: Jurnal Hukuk*, vol 6, no. 2, p. 231, 2023