



Implementasi Algoritma C4.5 Untuk Analisa Data Illegal Logging

Claudia Putri S.I. Butar-Butar

Program Studi Teknik Informatika, Universitas Budi Darma Medan, Indonesia

Email: Claudias.i.butarbutar@gmail.com

Abstrak—Data mining merupakan suatu penemuan informasi dengan melakukan penggalian pola informasi yang berisi pencarian trend dalam sejumlah data yang sangat besar membantu proses penyimpanan data dalam mengambil suatu keputusan di waktu yang akan datang. Dalam menentukan pola dilakukan teknik klasifikasi mengumpulkan *record (training data)*. Hutan adalah peranan penting yang sangat penting bagi pembangunan bangsa dan negara. Karena hutan dapat memberikan manfaat yang sebesar-besarnya. Namun, keadaan *Illegal Logging* saat ini telah mengalami penurunan luas yang drastis, untuk itu dengan memanfaatkan data *Illegal Logging* dapat menghasilkan informasi tentang *Illegal Logging* yang menjadi prioritas dan yang tidak prioritas untuk di tebang atau di tangani terlebih dahulu sehingga fungsi hutan benar pada penggunaannya. Algoritma C4.5 atau biasa di kenal dengan metode pohon keputusan dapat memberikan informasi *rule* data untuk menggambarkan proses yang terkait dengan pengolahan data *illegal logging*. Karakteristik data yang diklasifikasi dapat di peroleh dengan jelas, baik dalam bentuk struktur pohon keputusan maupun dalam bentuk aturan. Sehingga dalam tahap pengujian dengan software Weka dapat membantu dalam mengolah data *Illegal Logging* yang valid.

Kata Kunci: Dinas Kehutanan; Data Mining; Algoritma C4.5

Abstract—Data mining is an information discovery by extracting information partterns that contain trend searches in a large amount of data that helps the process of storing data in making a decision in the future. In determining then pattern classification techniques do to collect records (training data). Forests are an important role that is very important for nation and state development. Because forests can provide maximum benefits. However, that current state of illegal logging has experienced a drastic reduction in area, therefore utilizing illegal logging data can produce information about illegal logging that is a priority and which is not a priority and which is not a priority to be cut down or handled first so that the forest function is correct in its use. Algorithm C4.5 or commonly known as the decision tree method can provide rule data information to describe the processes associated with processing illegal logging data. The characteristic of the classified data can be obtained clearly, both in the form of decision tree structures and in the form of rules. So that in the testing phase with WEKA software can assist in processing valid illegal logging.

Keywords: Departement Forestry; Data Mining; Algorithm C4.5

1. PENDAHULUAN

Hutan merupakan paru-paru dunia yang dimana kawasan hutan ditumbuhi dengan pepohonan dan tumbuhan lainnya, sehingga berfungsi sebagai penampung karbon dioksida, habitat, hewan, serta pelestarian tanah, dan merupakan salah satu biosfer bumi yang terpenting, kita dapat menemukan hutan baik di daerah tropis maupun daerah beriklim dingin, di dataran rendah maupun di pegunungan di pulau kecil maupun di benua besar. Hutan merupakan suatu kumpulan tumbuhan dan juga tanaman, terutama pepohonan atau tumbuhan berkayu.

Analisa data illegal logging *Illegal Logging* merupakan suatu rangkaian kegiatan yang saling terkait mulai dari produsen kayu *illegal* yang melakukan penebangan kayu secara *illegal* hingga ke pengguna atau konsumen bahan baku kayu. Kayu tersebut akan melalui proses penyaringan yang *illegal*, pengangkutan *illegal* dan melalui proses penjualan yang *illegal*. *Illegal* merupakan suatu pelanggaran ataupun tidak ada izin dari pihak yang bersangkutan. *Logging* merupakan pemotongan, pemrosesan ditempat dan pemuatan pohon atau kayu[1].

Pada masa kini perkembangan teknologi komputer telah mengalami banyak perkembangan yang sangat pesat. Pemakaian teknologi komputer sebagai salah satu aplikasi dari teknologi informasi sudah menjadi suatu kebutuhan, karena banyak pekerjaan yang dapat diselesaikan dengan cepat. Komputer merupakan suatu alat elektronik yang mampu melakukan beberapa tugas antara lain menerima input data *illegal logging*, memproses input data *illegal logging* tadi sesuai dengan programnya, menyimpan perintah-perintah dan hasil dari pengolahan data *illegal logging* dan menyediakan output dalam bentuk informasi data *illegal logging*.

Algoritma C4.5 suatu algoritma yang digunakan unuk menghasilkan suatu pohon keputusan yang mampu mengklasifikasi suatu obyek. C4.5 merepresentasi konsep - konsep dalam bentuk pohon keputusan. Aturan-aturan yang dihasilkan oleh C4.5 mempunyai relasi yang hirarkis seperti suatu pohon (mempunyai akar, titik, cabang, dan daun). Beberapa menyebut struktur model yang dihasilkan C4.5 sebagai pohon keputusan (*decision tree*) sementara yang lain menyebutnya pohon aturan (*rule tree*). Dimana di dalam proses pengerjaan pada algoritma C4.5 menghitung nilai *gain* dan *entropy* terhadap atribut – atribut data yang telah disajikan oleh basis data sebelumnya. Dari proses klasifikasi dan hasil yang didapatkan digambarkan dalam bentuk pohon keputusan dan berdasarkan pohon keputusan tersebut didapatkanlah sebuah informasi baru yang berasal dari basis data sebelumnya dalam bentuk aturan – aturan atau *reale*.

Hasil dari penulisan ini dapat menghasilkan sebuah data yang akurat dalam pengolahan data *illegal logging*. Dapat disimpulkan bahwa dalam pengolahan *illegal logging* dapat diprediksi atau dapat dikelola data yang lama sampai data terbaru pengolahan hutan tersebut layak atau tidak dilindungi.



2. METODOLOGI PENELITIAN

Data mining, sering juga disebut *knowledge discovery in database* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam *set* data berukuran besar. Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan [1][2].

Dari defenisi-defenisi yang telah disampaikan, hal penting yang terkait dengan data mining adalah:

- Data mining merupakan suatu proses otomatis terhadap data yang sudah ada.
- Data yang akan diproses berupa data yang sangat besar.
- Tujuan data mining adalah mendapatkan hubungan atau pola yang akan mungkin memberikan indikasi yang bermanfaat[3].

Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewariskan banyak aspek dan teknik bidang-bidang ilmu yang sudah mapan terlebih dahulu. menunjukkan bahwa data mining memiliki akar panjang dari ilmu seperti ilmu kecerdasan buatan (*artificial intelligent*), *maching learning*, *statistic*, *database*, dan juga informasi *retrieve*[4][3].

2.1 Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk membentuk *decision tree* berdasarkan *training data*. Algoritma ini merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal [7]. *Decision tree* adalah model prediksi menggunakan struktur berhirarki[5]. Konsep dari *decision tree* adalah mengubah data menjadi *decision tree* dan aturan-aturan keputusan (*decision rules*). Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan C4.5 adalah bisa mengatasi *missing value*, *continue data* dan *pruning*. Algoritma C4.5 mempunyai input berupa *training samples* dan *samples*, *training samples* berupa data contoh yang akan digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya, sedangkan untuk *samples* merupakan *field-field* data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data [7][6]. Adapun langkah-langkah untuk penyelesaian Algoritma C4.5 yaitu sebagai berikut:

- Mencari Nilai *Entropy* dari kriteria-kriteria
- Mencari Nilai *Gain* dari setiap atribut
- Pembentukan Atribut sebagai akar berdasarkan *gain* tertinggi
- Pembentukan Cabang berdasarkan masing-masing nilai
- Ulangi Proses untuk masing-masing cabang [8].

3. HASIL DAN PEMBAHASAN

Pada pembahasan tentang menganalisa data *illegal logging* pada kantor dinas kehutanan provinsi sumatera utara, dengan menentukan beberapa atribut untuk mencapai hasil yang maksimal. Atribut merupakan kolom data yang menunjukkan fungsinya sebagai pembentuk karakteristik yang mempunyai nilai, kategori maupun kode untuk setiap basis data. Berikut yang merupakan keterangan atribut dalam pengolahan data *illegal logging* yaitu:

- Tanggal kejadian
Merupakan waktu kejadian pelanggaran yang dilakukan pada kawasan hutan.
- Jenis pelanggaran (Jl)
Kelas jenis laporan *illegal logging* ini melakukan pengecekan dominasi oleh adanya suatu jenis laporan Tebang, Bakar, tidak.
- Terlapor
Terlapor adalah orang yang telah melanggar undang –undang penyalagunaan hutan.
- Sanksi (S)
Kelas sanksi *illegal logging* diklasifikasikan menjadi kelas ada, tidak, ada.
- Barang Bukti (Bb)
Merupakan aktivitas yang sedang berlangsung yang turun ke lapangan pada *illegal logging* yang diklasifikasikan menjadi tidak,ada,tidak.
- Kawasan Hutan
Kawasan hutan ini merupakan daerah hutan yang telah terjadi pelanggaran *illegal logging*.
- Keterangan
Merupakan hasil penelitian bahwa keputusan memiliki kelas seperti penyidikan, Tersangka, Vonis.
Berikut adalah tabel *illegal logging* di Provinsi Sumatera Utara dapat dilihat pada tabel 1. dibawah ini:

Tabel 1. atribut Jumlah pelanggaran

Jumlah pelanggaran	Hukuman
1-5 kali	Ringan
5- 10 kali	Sedang
10-15kali	Berat

Berikut ini adalah tabel pembagian atribut tersangka dapat dilihat pada tabel berikut ini:

Tabel 2. atribut Jumlah tersangka

Jumlah Tersangka	Hukuman
1-5 kali	Ringan
5- 10 kali	Sedang
10-15kali	Berat

Selanjutnya pembagian atribut jlh tersangka dengan hukuman yang didapatkan dapat dilihat pada tabel berikut ini:

Tabel 3. atribut Jumlah saksi

Jumlah saksi	Hukuman
1-3	Ringan
3- 6	Sedang
6-9	Berat

Selanjutnya pembagian atribut Jlh saksi dengan hukuman yang didapatkan dapat dilihat pada tabel berikut ini:

Tabel 4. atribut Jumlah barang

Jumlah barang bukti	Hukuman
5 – 10	Ringan
10-15	Sedang
15- 20	Berat

Berdasarkan hasil pembagian data diatas, maka dilakukan penyusunan didalam tabel berikut ini:

Tabel 5. Data Illegal Logging

Lokasi kejadian	Tgl kejadian	Pelanggaran	Tersangka	Jlh saksi	Brg bukti	Hukuman
Desa Sibaganding	1 April 2016	15	11	9	5	Ringan
Gudang CV. Cipta Agro	22 Juni 2016	8	6	8	6	Sedang
Simpang Bukit	3 September 2016	7	5	7	7	Berat
UD. Mulia	15 September 2016	14	3	6	8	Ringan
Desa Sibaganding	14 Maret 2018	12	7	5	9	Sedang
Gudang CV. Cipta Agro	24 September 2016	13	9	4	10	Berat
Simpang Bukit	2 Nopember 2016	6	7	3		Sedang
UD. Mulia	20 Agustus 2018.	9	14	2	4	Sedang
Panglong Simbai	25 April 2018.	6	12	1	5	Berat
Desa Sibaganding	22 Mei 2017	5	12	8	3	Sedang
Gudang CV. Cipta Agro	11 Juni 2017.	4	8	4	11	Berat
Simpang Bukit	19 Juni 2017.	5	7	7	14	Ringan
UD. Mulia	8 Maret 2018	3	14	2	12	Berat
Panglong Simbai	13 Maret 2018	2	12	5	13	Ringan
Desa Sibaganding	22 Juni 2016	1	7	9	19	Sedang

Pilih Atribut Sebagai Akar

Untuk melihat atribut sebagai akar, didasarkan pada nilai *gain* tertinggi dari atribut-atribut yang ada dengan menggunakan dua persamaan maka akan didapatkan *entropy* dan *gain* yang digunakan sebagai akar dalam membuat pohon keputusan. Entropy total Untuk menggunakan 2 persamaan nilai entropy berdasarkan penilaian keseluruhan data berdasarkan tabel data decision illegal logic.

Keterangan :

1. Jumlah kasus (si) = 15
2. Nilai atribut (Ringan) = 4
3. Nilai atribut (Sedang) = 6
4. Nilai atribut (Berat) = 5



Dari keterangan di atas, dapat dihitung nilai entropy dengan rumus berikut ini :

$$\begin{aligned} \text{entropy (total}(s)) &= \left(-\frac{4}{15} * \log_2\left(\frac{4}{15}\right)\right) + \left(-\frac{6}{15} * \log_2\left(\frac{6}{15}\right) + \left(-\frac{5}{15} * \log_2\left(\frac{5}{15}\right)\right)\right) \\ &= (-4 : 15 * \text{LOG2}(4:15)) + (-6 : 15) * \text{LOG2}(6:15) + (-5:15) * \text{LOG2}(5:15)) \\ &= 1.671260397 \end{aligned}$$

Entropy (total) adalah menghitung nilai total hukuman ringan (4) dan sedang adalah (6) dan berat (5), sedangkan 15 adalah jumlah keseluruhan kasus.

1. Atribut Jlh Pelanggaran

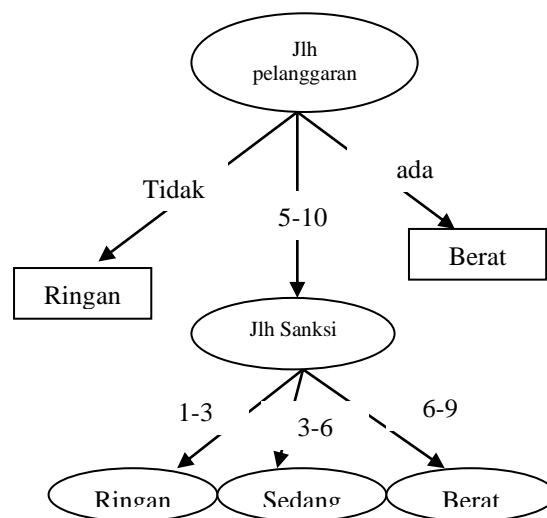
$$\begin{aligned} \text{entropy (5)} &= \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) \\ &= (-1 : 2 * \text{LOG2}(1:2)) + (-1 : 2) * \text{LOG2}(1:2) + (-0:2) * \text{LOG2}(0:2)) \\ &= 0 \\ \text{entropy (8)} &= \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) \\ &= (-0 : 1 * \text{LOG2}(0:1)) + (-1 : 1) * \text{LOG2}(1:1) + (-0:1) * \text{LOG2}(0:1)) \\ &= 0 \\ \text{entropy (7)} &= \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) \\ &= (-0 : 1 * \text{LOG2}(0:1)) + (-0 : 1) * \text{LOG2}(0:1) + (-1:1) * \text{LOG2}(1:1)) \\ &= 0 \\ \text{entropy (14)} &= \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) \\ &= (-1 : 1 * \text{LOG2}(1:1)) + (-0 : 1) * \text{LOG2}(0:1) + (-0:1) * \text{LOG2}(0:1)) \\ &= 0 \\ \text{entropy (12)} &= \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) \\ &= (-0 : 1 * \text{LOG2}(0:1)) + (-1 : 1) * \text{LOG2}(1:1) + (-0:1) * \text{LOG2}(0:1)) \\ &= 0 \\ \text{entropy (13)} &= \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) \\ &= (-0 : 1 * \text{LOG2}(0:1)) + (-0 : 1) * \text{LOG2}(0:1) + (-1:1) * \text{LOG2}(1:1)) \\ &= 0 \\ \text{entropy (6)} &= \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) \\ &= (-0 : 2 * \text{LOG2}(0:2)) + (-1 : 2) * \text{LOG2}(1:2) + (-1:2) * \text{LOG2}(1:2)) \\ &= 0 \\ \text{entropy (9)} &= \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) \\ &= (-0 : 1 * \text{LOG2}(0:1)) + (-1 : 1) * \text{LOG2}(1:1) + (-0:1) * \text{LOG2}(0:1)) \\ &= 0 \\ \text{entropy (6)} &= \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) \\ &= (-0 : 2 * \text{LOG2}(0:2)) + (-1 : 2) * \text{LOG2}(1:2) + (-1:2) * \text{LOG2}(1:2)) \\ &= 0 \\ \text{entropy (5)} &= \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) \\ &= (-1 : 2 * \text{LOG2}(1:2)) + (-1 : 2) * \text{LOG2}(1:2) + (-0:2) * \text{LOG2}(0:2)) \\ &= 0 \\ \text{entropy (4)} &= \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) \\ &= (-0 : 1 * \text{LOG2}(0:1)) + (-0 : 1) * \text{LOG2}(0:1) + (-1:1) * \text{LOG2}(1:1)) \\ &= 0 \\ \text{entropy (5)} &= \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) \end{aligned}$$

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n \frac{|jlh\ Saksi\ i|}{|total|} * entropy(jlh\ Saksi) \dots (6) \\
 &= 1.671260397 - \left(\left(\frac{2}{15} * 0 \right) + \left(\frac{2}{15} * 0 \right) + \left(\frac{2}{15} * 0 \right) + \left(\frac{1}{15} * 0 \right) + \left(\frac{2}{15} * 0 \right) + \left(\frac{2}{15} * 0 \right) + \left(\frac{1}{15} * 0 \right) \right. \\
 &\quad \left. + \left(\frac{2}{15} * 0 \right) + \left(\frac{1}{15} * 0 \right) \right) = 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 \\
 &= 0
 \end{aligned}$$

Tabel node 1

Tabel 6. Node

No	Total	Jlh kasus	Ringan	Sedang	Berat	entropy	Gain
1	Jlh pelanggaran	5	2	1	1	0	0
		8	1	0	1	0	0
		7	1	0	0	1	0
		14	1	1	0	0	0
		12	1	0	1	0	0
		13	1	0	0	1	0
		6	2	0	1	1	0
		9	1	0	1	1	0
		6	2	0	1	1	0
		5	2	1	1	0	0
		4	1	0	0	1	0
		5	2	1	1	0	0
		3	1	0	0	1	0
		2	1	1	0	0	0
		1	1	0	1	0	0
	Jlh sanksi	9	2	1	1	0	0
		8	2	0	2	0	0
		7	2	1	0	1	0
		6	1	1	0	0	0
		5	2	1	1	0	0
		4	2	0	0	2	0
		3	1	0	1	0	0
		2	2	0	1	1	0
		1	1	0	0	1	0



Gambar 1. Node Pohon Keputusan



Berdasarkan pembentukan pohon keputusan diatas, dan jumlah hasil perhitungan *entropy* dan *gainnya* sama dan mendapatkan nilai 0. Maka proses perhitungannya berhenti. Oleh karna itu maka didapatkan hasil analisa *illegal logging* yaitu jika jumlah pelanggaran 5 -10 dan jumlah saksi 6- 9 dinyatakan mendapatkan hukuman yang sangat berat.

4. KESIMPULAN

Berdasarkan hasil analisis dan implementasi yang dilakukan disimpulkan bahwa hasil penyelesaian *Algoritma C4.5* dapat mengetahui cara menganalisa data *illegal Logging*. Penggunaan *Algoritma C4.5* untuk perhitungan dalam mengidentifikasi faktor-faktor penyebab *illegal logging* ternyata mampu menyelesaikan permasalahan Dinas Kehutanan Sumatera Utara dalam mengelola data *illegal logging*.nn Pengujian yang dilakukan oleh data mining terhadap aplikasi *Weka* sangatlah mudah dalam mendapatkan informasi yang efektif serta hasil *rule* yang lebih mudah dipahami.

REFERENCES

- [1] N. Made, A. Santika, I. K. Gede, D. Putra, and I. M. Sukarsa, "Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan," vol. 6, no. 3, pp. 185–191, 2015.
- [2] H. Widayu, S. Darma, N. Silalahi, and Mesran, "Data Mining Untuk Memprediksi Jenis Transaksi Nasabah Pada Koperasi Simpan Pinjam Dengan Algoritma C4.5," *Issn 2548-8368*, vol. Vol 1, No, no. June, p. 7, 2017.
- [3] B. Santosa, *Data Mining, Teknik Pemanfaatan Data Untuk Keperluan Bisnis*, 10th ed. Yogyakarta: Graha Ilmu, 2007.
- [4] E. Ndruru and R. Limbong, "Implementasi Data Mining Dalam Pengelompokan Jurusan yang Diminati Siswa SMK Negeri 1 Lolowa menggunakan Metode Clustering," vol. 3, no. 2, pp. 107–113, 2018.
- [5] A. C. Wijaya, N. A. Hasibuan, and P. Ramadhani, "IMPLEMENTASI ALGORITMA C5 . 0 DALAM KLASIFIKASI PENDAPATAN MASYARAKAT (STUDI KASUS : KELURAHAN MESJID KECAMATAN MEDAN KOTA)," vol. 13, 2018.
- [6] Feri Sulianto dan Dominikus Juju, *data meramalkan bisnis perusahaan mining*, 19th ed. 2010.