

Data Mining Prediksi Besarnya Penggunaan Listrik Rumah Tangga di Kota Batam Dengan Menggunakan Algoritma C4.5

Yulia*, Nurul Azwanti

Universitas Putera Batam, Batam

*yulia_edwar@yahoo.co.id

Abstract

Human activity in the use of electricity has increased over time - time. This is due to electrical energy has become an important part for the development of human civilization in various fields including economic, technological, social and human culture. Strategy forecasting the need for electrical energy is needed. People's need for electric energy continues to grow every year. In addition to population growth, the economic growth of a region is believed to be one of the factors affecting the increasing consumption of electrical energy in the area. As the city of Batam in Batam Center area which is an industrial city and the population is fairly solid. Batam Center area includes the central area of Batam city because the area is close to Hang Nadim Airport Batam and Batam International Port Center. Therefore every household should understand the effective use of electricity so that the electricity needs do not become greater than the electricity supply. Data mining techniques with C4.5 algorithm can predict the use of household electricity to more easily regulate the use of household electricity. The sample data is taken as many as 30 correspondent data that use electricity meter in Batam Center area. The number of electronic goods, the number of users, the length of time at home and the area of the house building will be variable in analyzing the data. There are variables Wide of Home Build and Number of Family Members become decision forming tree variable. The calculation results have been tested using Weka 3.7.4 with the same rule result.

Keywords: *Algorithm C4.5; Data Mining; Electricity Usage; Prediction.*

Abstrak

Aktifitas manusia dalam penggunaan listrik mengalami peningkatan dari waktu – kewaktu. Hal ini diakibatkan karena energi listrik sudah menjadi bagian penting bagi perkembangan peradaban manusia di berbagai bidang antara lain bidang ekonomi, teknologi, sosial dan budaya manusia. Strategi prakiraan kebutuhan energi listrik sangat dibutuhkan. Kebutuhan masyarakat akan energi listrik terus tumbuh setiap tahunnya. Disamping pertumbuhan penduduk, pertumbuhan ekonomi suatu wilayah diyakini sebagai salah satu faktor yang mempengaruhi meningkatnya konsumsi energi listrik di daerah tersebut. Seperti kota Batam khususnya di kawasan Batam Center yang merupakan kota industri dan jumlah penduduknya yang terbilang padat. Wilayah Batam Center termasuk wilayah pusat kota Batam karena wilayah tersebut dekat dengan Bandara Hang Nadim Batam dan Pelabuhan Internasional Batam Center. Oleh karena itu setiap rumah tangga haruslah paham penggunaan listrik yang efektif sehingga kebutuhan listrik tidak menjadi lebih besar dari persediaan listrik. Teknik *data mining* dengan algoritma C4.5 mampu memprediksi penggunaan listrik rumah tangga agar lebih mudah mengatur penggunaan listrik rumah tangga. Data sampel yang diambil sebanyak 30 data koresponden yang menggunakan listrik meteran di kawasan Batam Center. Jumlah barang elektronik, jumlah pemakai, lama waktu di rumah dan luas bangunan rumah akan menjadi variabel dalam menganalisa data. Didapat variabel Luas Bangun Rumah dan Jumlah Anggota Keluarga menjadi variabel pembentuk pohon keputusan. Hasil perhitungan telah diuji dengan menggunakan Weka 3.7.4 dengan hasil *rule* yang sama.

Kata Kunci: *Algoritma C4.5; Data Mining; Pemakaian Listrik; Prediksi.*

1. Pendahuluan

Listrik telah menjadi kebutuhan yang tidak terlepas dari kegiatan yang dilakukan oleh manusia sehari-hari. Peranan listrik tidak hanya menjadi kebutuhan sekunder, namun telah beralih menjadi kebutuhan primer dikarenakan tanpa adanya listrik segala kegiatan yang dilakukan akan sangat menghambat. Mulai dari peralatan rumah tangga seperti kulkas, kipas angin, setrika, *mixer*, *rice cooker* hingga alat komunikasi seperti *handphone*, laptop dan berbagai macam alat elektronik lainnya sangat membutuhkan adanya energi listrik. Perkembangan teknologi saat ini juga mempengaruhi dari penggunaan listrik, biasanya semakin banyak peralatan baik teknologi maupun elektronik sering kali menjadi tolak ukur taraf kesejahteraan dari masyarakat (Rahman & Nanggalo, 2015). Seperti yang kita ketahui, semakin banyak penggunaan listrik, maka semakin banyak pula energi listrik yang dibutuhkan.

Berdasarkan data Badan Pusat Statistik Kepulauan Riau pada tahun 2016 jumlah penduduk di Kota Batam mencapai 1.236.399 jiwa yang tersebar di berbagai daerah di Kota Batam. Wilayah Batam Center termasuk wilayah pusat kota Batam karena wilayah tersebut berdekatan dengan Bandara Hang Nadim Batam dan Pelabuhan Internasional Batam Center. Kepadatan penduduk tersebut juga menyebabkan kebutuhan akan energi listrik menjadi sangat tinggi. Jika energi listrik terus menerus dikonsumsi maka bisa menyebabkan habisnya persediaan tenaga listrik. Hal inilah yang perlu disadari oleh masyarakat perlunya melakukan penghematan listrik untuk keseimbangan dari pasokan tenaga listrik yang tersedia, terlebih lagi terhadap kesadaran dalam pemakaian listrik rumah tangga (Sari, Moh. Djemdjem Djamaludin, & Anggi Mayang, 2011).

Untuk mengetahui penggunaan listrik rumah tangga diperlukan adanya prediksi. Prediksi penggunaan listrik rumah tangga dapat menggunakan teknik *data mining*. *Data mining* bekerja dengan melakukan proses penggalian informasi dari data penggunaan listrik maupun dari basis data (Saleh, 2015). Teknik dalam melakukan prediksi dapat menggunakan teknik klasifikasi pada Algoritma C4.5. Algoritma C4.5 memiliki banyak kelebihan diantaranya dapat menghasilkan model berupa pohon atau aturan yang dapat dengan mudah diinterpretasikan. Pohon atau aturan yang terbentuk dapat membantu dalam membaca prediksi yang akan diterapkan pada prediksi penggunaan listrik rumah tangga.

2. Kajian Literatur

2.1 Data Mining

Data Mining disebut juga *Knowledge Discovery in Database (KDD)* melakukan ekstraksi informasi dari tumpukan data. Proses pencarian informasi tersebut akan menemukan pola kecenderungan dari data yang kemudian hasil dari penambangan (*mining*) dapat menjadi informasi yang mudah dipahami (Siburian, 2014). *Data Mining* mengacu pada proses pencarian informasi yang tidak diketahui sebelumnya dari sekumpulan data besar (Selvia Lorena Br Ginting, Wendi Zarman, 2014).

2.2 Pohon Keputusan (Decision Tree)

Pohon keputusan menggunakan representasi struktur pohon (*tree*) di mana setiap *node* merepresentasikan atribut, cabangnya merepresentasikan nilai dari atribut dan daun merepresentasikan kelas.

2.3 Algoritma C4.5

Untuk membentuk pohon keputusan, maka dibutuhkan Algoritma C4.5 yang berguna sebagai algoritma untuk mengetahui keputusan dari suatu kasus. Algoritma C.45 adalah salah satu algoritma induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel *training*, label *training* dan atribut. Algoritma C4.5 merupakan pengembangan dari ID3. Beberapa pengembangan yang dilakukan pada C4.5 adalah sebagai antara lain bisa mengatasi *missing value*, bisa mengatasi *continua data*, dan *pruning* (Faradillah, 2013).

Dalam pembentukan pohon keputusan, untuk memilih atribut akar, didasarkan pada *nilai gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti yang tertera dalam persamaan berikut.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Rumus 1. Mengitung Gain

Di mana :

- S : himpunan kasus
- A : atribut
- N : jumlah partisi atribut A
- |S_i| : jumlah kasus pada partisi ke-*i*
- |S| : jumlah kasus dalam S

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan 2 berikut:

$$Entropy(S) = \sum_{i=1}^n -pi * \log_2 pi$$

Rumus 2. Mengitung *Entropy*

Di mana:

- S : himpunan kasus
- A : fitur
- N : jumlah partisi S
- pi : proporsi dari Si terhadap S

2.4 Tinjauan Umum Listrik Rumah Tangga

Menurut (Widjayanti, 2007) untuk penggunaan energi akhir bagi peralatan listrik baik untuk menggerakkan motor, lampu penerangan, memanaskan, mendinginkan ataupun untuk menggerakkan kembali suatu peralatan mekanik untuk menghasilkan bentuk energi maka dibutuhkan adanya energi listrik. Dalam halnya rumah tangga, contoh pemakaian energi listrik seperti mesin cuci, kipas angin, kulkas, setrika dan lain sebagainya. Pemakaian listrik ini juga dilatarbelakangi dari kemajuan teknologi yang mempermudah manusia dalam melakukan pekerjaannya. Bisa kita bayangkan jika saat ini kita belum pernah sama sekali menyentuh fungsi dari listrik. Namun, dibalik penggunaannya yang sangat membantu, konsumsi energi listrik juga harus diseimbangkan mengingat cadangan energi listrik akan menipis seiring dengan penggunaannya secara terus menerus.

2.5 Penelitian Terdahulu

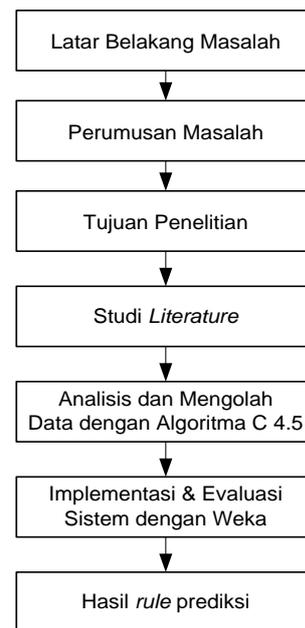
Tabel 1. Penelitian Terdahulu

Judul Penelitian	Hasil Penelitian
Implementasi Metode Klasifikasi <i>Naive Bayes</i> Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga (Saleh, 2015)	Metode <i>Naive Bayes</i> memanfaatkan data <i>training</i> untuk menghasilkan probabilitas setiap kriteria untuk <i>class</i> yang berbeda, mengklasifikasikan 47 data dari 60 data yang diuji. Sehingga metode <i>Naive Bayes</i> berhasil memprediksi besarnya penggunaan listrik rumah tangga dengan persentase keakuratan sebesar 78,3333%.
<i>Data Mining</i> Peramalan Konsumsi Listrik dengan Pendekatan <i>Cluster Time Series</i> sebagai <i>Preprocessing</i> (Riyadi & Fithriasari, 2016)	Hasil yang diperoleh adalah dihasilkan sebanyak tujuh <i>cluster</i> dengan anggota <i>cluster</i> terbanyak pada <i>cluster</i> ke empat yakni sebanyak 120 <i>client</i> . Selanjutnya model peramalan dengan menggunakan ANN lebih baik dibandingkan ARIMA. Diperoleh sebanyak 259 dari 348 <i>client</i> yang menyatakan bahwa permodelan dengan menggunakan ANN lebih baik dibandingkan ARIMA.

Penentuan Daya Listrik Rumah Tangga Menggunakan Metode *Decision Tree* (Tanjung, Sentinuwo, & Jacobus, n.d.) Penggunaan algoritma dari metode *decision tree* baik algoritma ID3 maupun C4.5 dapat menentukan daya listrik rumah tangga. Hasil akurasi kedua algoritma menggunakan *confusion matrix* dengan data sejumlah 300 dan 150 data *testing* menunjukkan algoritma C4.5 dengan akurasi 88 % lebih akurat daripada algoritma ID3 yang memiliki akurasi 62 %.

3. Metode Penelitian

Metode penelitian dapat digambarkan pada gambar 1 berikut ini:



Gambar 1. Metode Penelitian

Keterangan dari gambar 1 di atas diuraikan sebagai berikut:

- (1) Latar belakang masalah
Tahap ini merupakan langkah awal dalam penelitian yaitu membahas suatu sistem yang melatar balakangi dalam suatu permasalahan pokok yang paling mendasar dalam penelitian untuk menggali permasalahan yang ada.
- (2) Perumusan Masalah
Pada tahap ini akan merumuskan masalah-masalah yang terjadi pada sistem sehingga pada tahap ini akan dilakukan proses pengamatan, pendefinisian dan *explorasi* yang nantinya akan diteliti lebih lanjut.
- (3) Tujuan Penelitian
Dari perumusan masalah yang ada, maka tahap penentuan tujuan berguna untuk memperjelas sasaran dari penelitian. Adapun tujuan dari penelitian ini yaitu memprediksi besarnya penggunaan listrik rumah tangga.

- (4) Studi *Literature*
Studi *literature* yaitu melakukan studi pustaka dengan tujuan untuk mencari dasar-dasar referensi yang kuat bagi peneliti dalam menerapkan metode yang tepat dalam penelitian.
- (5) Analisis dan Mengolah Data dengan Algoritma C 4.5
Dalam penelitian ini data penggunaan listrik akan diklasifikasi atau dikelompokkan dengan menggunakan teknik Algoritma C4.5.
- (6) Implementasi dan Evaluasi Sistem dengan WEKA
Pada tahap ini dilakukan implementasi dan evaluasi terhadap sistem dan *knowledge* yang dihasilkan pada tahap analisa dan rancangan sistem. Untuk memastikan bahwa hasil yang diperoleh pada tahap sebelumnya telah bebas dari kesalahan, maka dilakukan implementasi sistem dengan menggunakan aplikasi WEKA.
- (7) Hasil *Rule* Prediksi
Pada tahap ini dibandingkan hasil *rule* prediksi dari pohon keputusan yang terbentuk baik dari hitungan manual maupun dari pengujian dengan WEKA.

4. Hasil dan Pembahasan

4.1 Hasil Penelitian

Hasil data yang telah dikumpulkan dilakukan klasifikasi guna mempermudah dalam melakukan perhitungan dengan algoritma C4.5

Tabel 2. Klasifikasi Total Barang Elektronik

Total Barang	Klasifikasi
1-11	Sedikit (S)
>12	Banyak (B)

Tabel 3. Klasifikasi Jumlah Anggota Keluarga

Jumlah Anggota	Klasifikasi
1-3	Sedikit (S)
>3	Banyak (B)

Tabel 4. Klasifikasi Lama Waktu di Rumah

Lama Waktu	Klasifikasi
1-12	Sebentar (S)
>12	Lama (L)

Tabel 5. Klasifikasi Luas Bangun Rumah

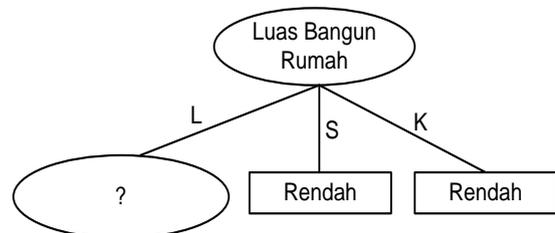
Type Rumah	Klasifikasi
30	Kecil (K)
32	Sedang (S)
36	Luas (L)

Dari data klasifikasi, maka didapat data perhitungan sebagai berikut:

Node	Jumlah Kasus(S)	Ya(S1)	Tidak (S2)	Entropy	Gain
1 Total		30	16	14	0.99679
Total Barang Elektronik					
Banyak (B)	21	12	9	0.98523	0.00981
Sedikit (S)	9	4	5	0.99108	
Jumlah Anggota Keluarga					
Banyak (B)	15	10	5	0.9183	0.05217
Sedikit (S)	15	6	9	0.97095	
Lama Waktu di Rumah					
Lama (L)	27	15	12	0.99108	0.01299
Sebentar (S)	3	1	2	0.9183	
Luas Bangun Rumah					
Luas (L)	21	14	7	0.77995	0.17534
Sedang (S)	6	2	4	0.9183	
Kecil (K)	3	1	2	0.9183	

Gambar 2. Perhitungan Node 1

Hasil perhitungan yang terdapat pada gambar 2 dapat dilihat bahwa Gain tertinggi 0,17534 terletak pada variabel Luas Bangun Rumah dimana Luas Bangun Rumah S dengan keputusan Rendah, Luas Bangun Rumah K dengan keputusan Rendah, sedangkan Luas Bangun Rumah L harus dicari lagi perhitungannya. Gambar pohon keputusan yang terbentuk dapat dilihat pada gambar 3 berikut.



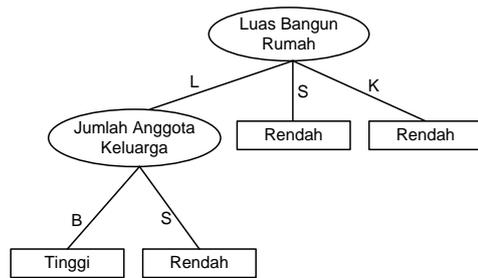
Gambar 3. Pohon Keputusan Node 1

Dilakukan lagi perhitungan terhadap Node 1.1 dengan menyortir data hanya pada kasus Luas Bangun Rumah L. hasil perhitungan dapat dilihat pada gambar 4.

Node	Jumlah Kasus(S)	Ya(S1)	Tidak (S2)	Entropy	Gain
1.1 Luas Bangun Rumah L		21	14	7	0.9182958
Total Barang Elektronik					
Banyak (B)	14	11	3	0.7495953	0.417094
Sedikit (S)	7	3	4	0.9852281	
Jumlah Anggota Keluarga					
Banyak (B)	9	9	0	0	0.604844
Sedikit (S)	12	5	7	0.9798688	
Lama Waktu di Rumah					
Lama (L)	18	13	5	0.8524052	0.393519
Sebentar (S)	3	1	2	0.9182958	

Gambar 4. Perhitungan Node 1.1

Hasil perhitungan yang terdapat pada gambar 4 didapat lagi Gain tertinggi 0,604844 terletak pada variabel Jumlah Anggota Keluarga dimana Jumlah Anggota Keluarga B dengan keputusan Tinggi dan Jumlah Anggota Keluarga S dengan keputusan Rendah. Gambar pohon keputusan yang terbentuk dapat dilihat pada gambar 5 berikut.



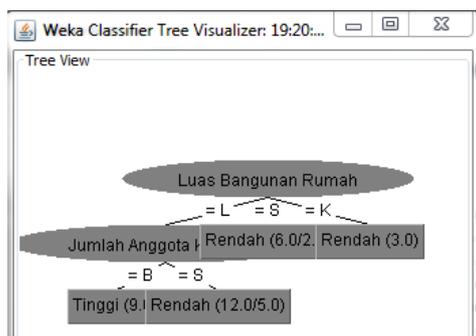
Gambar 5. Pohon Keputusan Akhir

4.2 Pengujian

Pengujian dari perhitungan yang telah dilakukan dapat menggunakan *software* WEKA. Adapun langkah-langkah dalam mengolah data menggunakan *WEKA* adalah sebagai berikut:

- (1) Seluruh data variabel dan keputusan yang akan diproses oleh *WEKA* disimpan terlebih dahulu di *Microsoft Excel* dengan format *.xls*. Kemudian ubah format data tersebut dan *Save as type* cari format *.csv*;
- (2) Buka *software WEKA 3.8.1*, klik dua kali di *shortcutnya* atau cari melalui *My Computer* dan akan muncul tampilan *WEKA*;
- (3) Pilih *Explorer*, kemudian pilih *Open file*, cari dimana *file* penjualan.csv tersebut berada, pilih dan klik *Open*;
- (4) Selanjutnya pilih variabel yang mempengaruhi data yang akan diproses;
- (5) Klik menu *Classify*, pada *Classifier* klik *Choose*, untuk algoritma C4.5 pilih *trees* dan klik *J48*;
- (6) Pilih *Use training set* lalu klik tombol *Start*. Akan tampil *Classifier output* di area sebelah kanan yang merupakan hasil pengolahan dari *WEKA*. *Classifier output* berisi *running* data yang diinputkan yang merupakan atribut-atribut yang akan membentuk pohon keputusan. Informasi yang ditampilkan berupa jumlah kasus dan keputusannya dan banyaknya cabang dari pohon keputusan;
- (7) Untuk melihat pohon keputusan yang terbentuk dari hasil pengolahan *WEKA* adalah klik kanan pada *Result list*, pilih *Visualize tree*.

Hasil pengujian dapat dilihat pada gambar 6.



Gambar 6. Hasil Pengujian dengan WEKA

4.3 Pembahasan

Dari hasil perhitungan manual yang dilakukan *rule* yang terbentuk sebagai berikut:

- (1) Jika luas bangun rumah kecil, maka penggunaan listrik rendah.
- (2) Jika luas bangun rumah sedang, maka penggunaan listrik rendah.
- (3) Jika luas bangun rumah luas dan jumlah anggota keluarga banyak, maka penggunaan listrik tinggi.
- (4) Jika luas bangun rumah luas dan jumlah anggota keluarga sedikit, maka penggunaan listrik rendah.

Kemudian dibandingkan dengan hasil pengujian menggunakan *WEKA* maka didapat *rule* prediksi besarnya penggunaan listrik rumah tangga adalah sebagai berikut:

- (1) Jika luas bangun rumah kecil, maka penggunaan listrik rendah.
- (2) Jika luas bangun rumah sedang, maka penggunaan listrik rendah.
- (3) Jika luas bangun rumah luas dan jumlah anggota keluarga banyak, maka penggunaan listrik tinggi.
- (4) Jika luas bangun rumah luas dan jumlah anggota keluarga sedikit, maka penggunaan listrik rendah.
- (5) Proses secara manual dan menggunakan *software WEKA* membentuk pohon keputusan dan *rule* yang sama sehingga dipastikan pengujian dilakukan dengan sangat baik.

5. Kesimpulan dan Saran

Kesimpulan

Kesimpulan dari penelitian ini adalah sebagai berikut: (1) Algoritma C4.5 dapat digunakan sebagai metode klasifikasi dalam memprediksi besarnya penggunaan listrik di Kota Batam dengan memperhatikan nilai *gain* (penguatan) tertinggi dari empat atribut seperti jumlah barang elektronik, jumlah pemakai, lama waktu di rumah dan luas bangunan rumah. (2) Hasil perhitungan dengan Algoritma C4.5 dengan metode pohon keputusan dapat memberikan informasi *rule* prediksi untuk menggambarkan proses yang terkait dengan prediksi besarnya penggunaan listrik rumah tangga dan (3) Dari keempat variabel yang diolah didapat bahwa variabel luas bangunan rumah dan banyaknya jumlah anggota keluarga di rumah yang memakai daya listrik sangat mempengaruhi besarnya penggunaan listrik rumah tangga.

Saran

Saran dalam penelitian ini adalah (1) Penambahan variabel sangat diperlukan guna melihat variabel manakah yang paling mempengaruhi dari penggunaan listrik rumah

tangga. (2) Untuk mendapatkan hasil prediksi yang lebih baik lagi, perlu menggabungkan atau membandingkan beberapa metode klasifikasi yang lain.

Ucapan Terima Kasih

Ucapan terima kasih kepada berbagai pihak yang telah mendukung penelitian ini, diantaranya DIKTI yang telah memberikan dana penelitian, Rektor Universitas Putera Batam Ibu Dr. Nur Elfi Husda, S. Kom., M. SI atas dukungannya, Ibu Nia Ekawati, S. Kom., M. SI sebagai ketua LPPM atas dukungannya dan para responden yang telah bekerja sama dalam mengisi isian data penelitian yang peneliti berikan.

Daftar Pustaka

- Faradillah, S. (2013). Implementasi Data Mining Untuk Pengenalan Karakteristik Transaksi Customer Dengan Menggunakan Algoritma C4.5, 63–70.
- Rahman, A., & Nanggalo, K. (2015). Prakiraan Dan Analisa Kebutuhan Energi Listrik Provinsi Sumatera Barat Hingga Tahun 2024 Dengan Metode Analisis Regresi Linear Berganda. *Jurnal Teknik Elektro ITP*, 4(2).
- Riyadi, M. A. A., & Fithriasari, K. (2016). Data Mining Peramalan Konsumsi Listrik dengan Pendekatan Cluster Time Series sebagai Preprocessing, 5(1).
- Saleh, A. (2015). Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Journal*, 2, 207–217.
- Sari, L. N. Y., Moh. Djemdjem Djamaludin, & Anggi Mayang. (2011). Analisis Sikap dan Perilaku Penghematan Listrik Pada Sektor Rumah Tangga, 4(1), 82–90.
- Selvia Lorena Br Ginting, Wendi Zarman, I. H. (2014). Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik, (November).
- Siburian, B. R. (2014). Aplikasi Data Mining Untuk Menampilkan Tingkat Kelulusan Mahasiswa Dengan Algoritma Apriori. *Pelita Informatika Budi Darma*, VII, 56–61.
- Tanjung, Y. P., Sentinuwo, S., & Jacobus, A. (n.d.). Penentuan Daya Listrik Rumah Tangga Menggunakan Metode Decision Tree.
- Widjayanti, W. (2007). Profil Konsumsi Energi Listrik Pada Hunian Rumah Tinggal Studi Kasus Rumah Desain Minimalis Ditinjau Dari Aspek Pencahayaan Buatan. *Enclosure*, 6(2), 97–106.