

Klasifikasi User Berdasarkan Trafik Http/Https Menggunakan Metode Naïve Bayes

Eko Prayoga¹, T.M Diansyah², Risiko Liza³

^{1,2,3}Program Studi Teknik Informatika Fakultas Teknik dan Komputer
Universitas Harapan Medan
Jl. H.M. Jhoni No 70 Medan, Indonesia
Email: ekoprayoga18@gmail.com

Abstract

Along with the times and accompanied by advances in information and communication technology, it is undeniable that at this time all activities use information technology. One of the activities is to use the internet. The research will conduct a classification based on internet usage data obtained through questionnaires using data mining techniques. The attributes that will be used in doing the classification are Name, Age, Gender, Last Education. The method used is the Naïve Bayes method, which is one of the classification techniques in data mining. Based on the research conducted, it was concluded that based on internet user data used as training data, the Naïve Bayes method succeeded in classifying 32 data from 50 data tested. So the Naïve Bayes method succeeded in predicting the magnitude of the percentage of accuracy by 64%.

Keywords : *Data Mining, Classification, Naïve Bayes*

Abstrak

Seiring dengan perkembangan zaman dan diiringi dengan kemajuan teknologi informasi dan komunikasi, maka tidak bisa dipungkiri jika pada saat ini segala kegiatan menggunakan teknologi informasi. Salah satu kegiatannya adalah dengan menggunakan internet. Penelitian akan melakukan klasifikasi berdasarkan data penggunaan internet yang diperoleh melalui kuesioner dengan menggunakan teknik data mining. Atribut yang akan digunakan dalam melakukan klasifikasi adalah Nama, Usia, Jenis Kelamin, Pendidikan Terakhir. Metode yang digunakan adalah metode *Naïve Bayes*, yang merupakan salah satu teknik pengklasifikasian dalam data mining. Berdasarkan penelitian yang dilakukan dihasilkan kesimpulan bahwa Berdasarkan data pengguna internet yang dijadikan *data training*, metode *Naïve Bayes* berhasil mengklasifikasikan 32 data dari 50 data yang diuji. Sehingga metode *Naïve Bayes* berhasil memprediksi besarnya presentase keakuratan sebesar 64%.

Kata kunci : *Data Mining, Klasifikasi, Naïve Bayes*

1. PENDAHULUAN

Seiring dengan perkembangan zaman dan diiringi dengan kemajuan teknologi informasi dan komunikasi, maka tidak bisa dipungkiri jika pada saat ini segala kegiatan menggunakan teknologi informasi. Salah satu kegiatannya adalah dengan menggunakan internet. Internet merupakan salah satu produk dari teknologi informasi dan komunikasi yang tingkat penggunaannya semakin meningkat, mulai dari anak-anak, remaja, dewasa, laki-laki, perempuan, miskin dan kaya. Internet mempengaruhi hampir dalam semua aspek kehidupan manusia. Media seperti surat kabar, majalah, radio, dan televisi merupakan sumber informasi yang paling efektif. Realitas saat ini, media seperti facebook, youtube, twitter, instagram, whatsapp, dan sejenisnya menjadi media komunikasi dan informasi yang paling digemari. Protokol transfer utama yang digunakan oleh WWW adalah HTTP. Ini adalah protokol aplikasi berbasis klien server sederhana yang dibangun di atas TCP (*Transmission Control Protocol*). Transaksi yang

khas dari HTTP ini adalah klien melakukan koneksi dengan server HTTP, dan melakukan permintaan untuk sumber yang diinginkan dan menunggu respon dari server. Setelah menerima permintaan dari klien, server akan memproses permintaan klien dan mengirimkan respon dan menutup koneksi. Respon dapat berupa sumber yang diinginkan atau dapat juga hasil dari eksekusi dalam kasus *resource* yang diminta adalah sumber yang dapat dieksekusi. Klien mengalamatkan sumber yang dibutuhkan dengan URL (*Uniform Resource Locator*), yaitu alamat fisik dari objek yang dapat diambil menggunakan protokol jaringan seperti HTTP. URL terdiri atas nama server (DNS atau alamat IP) dan nama dari sumber yang diminta. Nama server (DNS) dan nama sumber dipisah dengan tanda “/”. Sebagai contoh di bawah ini menunjukkan permintaan HTTP yang berisi URL dari sumber yang diminta. <http://erma.akprind.ac.id/~erma/index.html>; erma.akprind.ac.id adalah nama DNS dari server, dan ~erma/index.html adalah nama dari sumber.

Ada beberapa operasi HTTP yang salah satu contohnya adalah GET dan POST. Operasi GET digunakan untuk meminta sumber dari server. Operasi POST digunakan untuk mengirim informasi ke sumber yang ada di server. *Bayesian classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* [1].

2. METODOLOGI PENELITIAN

Adapun metode penelitian yang digunakan dalam penelitian ini antara lain :

a) Studi Literatur

Studi literatur dimaksud untuk mengumpulkan data dari berbagai sumber seperti buku, jurnal ilmiah dan lain sebagainya yang membahas tentang klasifikasi menggunakan metode *naïve bayes*. Kemudian mempelajari data yang dianggap penting dan ada kaitannya dengan penulisan tugas akhir ini untuk menguatkan ide dan pemikiran penulis

b) Analisa Kebutuhan Sistem

Menganalisis kebutuhan yang diperlukan untuk mengembangkan dan mengimplementasikan pengklasifikasian user berdasarkan trafik HTTP/HTTPS.

c) Implementasi Perancangan Sistem

Pada tahap ini dilakukan perancangan klasifikasi user berdasarkan trafik HTTP/HTTPS menggunakan metode *naïve bayes*, sehingga dapat berfungsi untuk mengklasifikasi user berdasarkan trafik HTTP/HTTPS menggunakan metode *naïve bayes*.

d) Pengujian dan Hasil

Dilakukan pengujian terhadap keberhasilan proses perancangan klasifikasi user berdasarkan trafik HTTP/HTTPS menggunakan metode *naïve bayes*. Dimana dalam perancangan tersebut berjalan dengan baik.

3. HASIL DAN PEMBAHASAN

a) Data Mining

Data mining merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database*, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar. analisa terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut. Data mining bukanlah suatu bidang yang sama sekali baru. Salah satu kesulitan untuk mendefinisikan data mining adalah kenyataan bahwa data mining mewarisi banyak aspek dan teknik dari

bidang-bidang ilmu yang dulu sudah mapan terlebih dulumenunjukkan bahwa data mining memiliki akar yang Panjang dari bidang ilmu yang berbeda seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, *statistik*, *database*, dan juga *information retrieval* [2].

b) Naïve Bayes

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

Metode Bayes merupakan pendekatan statistic untuk melakukan inferensi induksi pada persoalan klasifikasi. Pertama kali dibahas terlebih dahulu tentang konsep dasar dan definisi pada Teorema Bayes, kemudian menggunakan teorema ini untuk melakukan klasifikasi dalam Data Mining.

c) Klasifikasi

Klasifikasi merupakan teknik dalam data mining untuk mengelompokkan data berdasarkan keterikatan data terhadap data sampel. Dalam menganalisa performa dari beberapa teknik klasifikasi, maka dilakukan perbandingan metode dalam data mining baik metode decision tree maupun metode machine learning untuk memilih metode yang terbaik dengan akurasi yang tinggi dalam mengklasifikasi dataset Bank Direct Marketing. Menggunakan model klasifikasi yang ada pada software opensource dalam data mining yaitu berupa constant, adaboost, stochastic gradient descent, k-nn, cn-2 rule inducer, svm, naive bayes, random forest dan tree.

Untuk klasifikasi, ketika ada dua atau lebih kelas mayoritas, classifier memilih kelas prediksi secara acak, tetapi selalu mengembalikan kelas yang sama untuk contoh tertentu[3].

d) Pengguna Internet

Penggunaan internet sebagai media pembelajaran dapat dianggap sebagai suatu hal yang sudah jamak digunakan dikalangan pelajar. Dengan adanya internet sebagai sumber belajar memudahkan kita untuk mengakses berbagai sumber informasi yang tersedia, karena internet dapat membantu kita meningkatkan taraf hidup melalui pendidikan. Internet juga dapat mengakses berbagai referensi, baik yang berupa hasil penelitian, maupun artikel hasil kajian dalam berbagai bidang. Informasi yang tersedia dan dapat diakses melalui internet terjadi di seluruh dunia (global world). Pesatnya perkembangan internet baik di negara maju maupun di negara yang sedang berkembang salah satu teknologi informasi yang berkembang adalah internet [4].

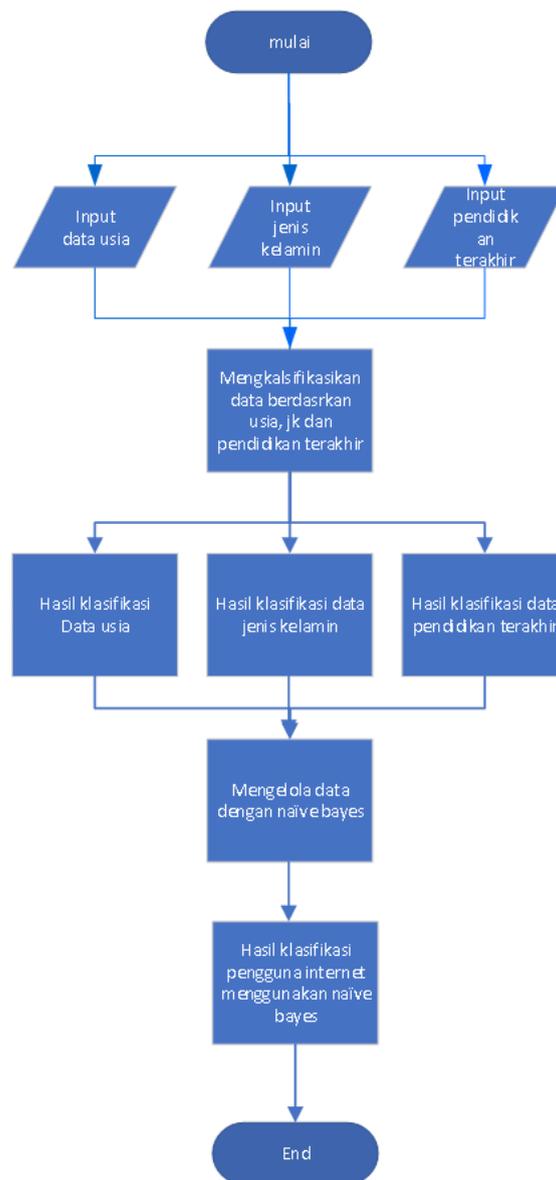
e) HTTP / HTTPS

HTTP adalah sebuah protokol meminta atau menjawab antara client dan server. Sebuah client HTTP seperti web browser, biasanya memulai permintaan dengan membuat hubungan TCP/IP ke port tertentu di tuan rumah yang jauh (biasanya port 80). Sebuah server HTTP yang mendengarkan di port tersebut menunggu client mengirim kode permintaan (request), seperti "GET / HTTP/1.1" (yang akan meminta halaman yang sudah ditentukan), diikuti dengan pesan MIME yang memiliki beberapa informasi kode kepala yang menjelaskan aspek dari permintaan tersebut, diikuti dengan badan dari data tertentu. Beberapa kepala (header) juga bebas ditulis atau tidak, sementara lainnya

(seperti tuan rumah) diperlukan oleh protokol HTTP/1.1. Begitu menerima kode permintaan (dan pesan, bila ada), server mengirim kembali kode jawaban, seperti "200 OK", dan sebuah pesan yang diminta, atau sebuah pesan error atau pesan lainnya. Pengembangan HTTP dikoordinasi oleh Konsorsium World Wide Web (W3C) dan grup bekerja Internet Engineering Task Force (IETF), bekerja dalam publikasi satu seri RFC, yang paling terkenal RFC 2616, yang menjelaskan HTTP/1.1, versi HTTP yang digunakan umum sekarang ini.

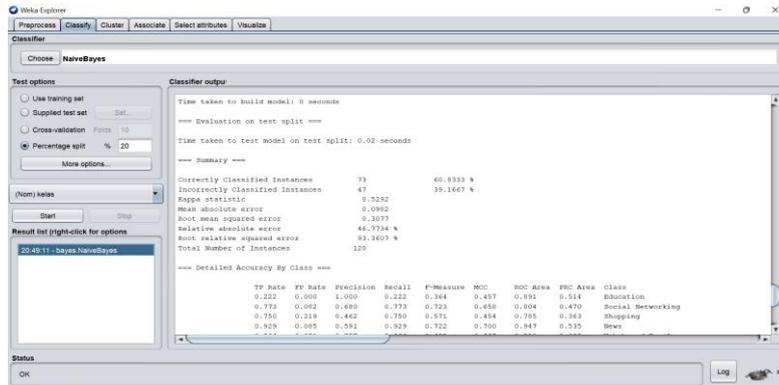
HTTPS dienkripsi dan deskripsi dari halaman yang di minta oleh pengguna dan halaman yang di kembalikan oleh web server. Kedua protokol tersebut memberikan perlindungan yang memadai dari serangan eavesdroppers, dan man in the middle attacks. Pada umumnya port yang digunakan HTTPS adalah port 443. Tingkat keamanan tergantung pada ketepatan dalam mengimplementasikan pada browser web dan perangkat lunak server dan didukung oleh algoritma penyandian yang aktual. Oleh karena itu, pada halaman web digunakan HTTPS, dan URL yang digunakan dimulai dengan https://.

Sistem pada klasifikasi user menggunakan *naïve bayes* dirancang dengan diagram alir flowchart yang menjelaskan tentang alur pengguna internet yang nantinya dibedakan oleh beberapa kategori. Hasil ini akan membuat sebuah analisis yang lebih baik dalam segi hasil pembahasan diagram flowchart dapat dilihat dibawah ini :



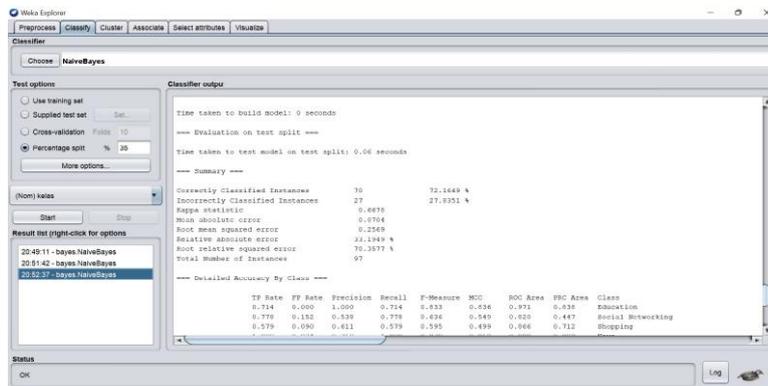
Gambar 1. Diagram Klasifikasi User Menggunakan *Naïve Bayes*

Berikut ini akan di tampilkan hasil dari klasifikasi user menggunakan aplikasi weka yang digunakan untuk mendeskripsikan tentang halaman-halaman yang ada pada klasifikasi user berdasarkan trafik HTTP/HTTPS menggunakan *Naive Bayes*. Hasil penerapannya dapat dilihat dan akan dijelaskan pada setiap tampilan yang ada pada aplikasi. Analisis data Penggunaan internet menggunakan *Naïve Bayes* dengan metode *test percentage split* 20% maksudnya ialah 20% dari data akan digunakan sebagai data *training* sementara 80% dari data akan digunakan sebagai data *testing*. Hasil aplikasi *Weka* dalam mengklasifikasi user dengan menggunakan metode *test percentage split* 20% adalah sebagai berikut :



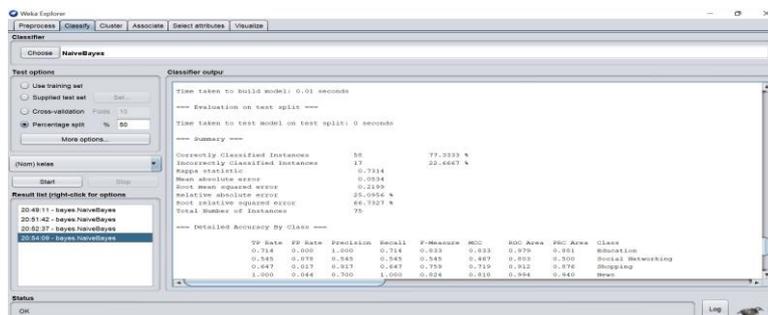
Gambar 2. Tampilan Form Hasil Analisis 20%

Analisis data Penggunaan internet menggunakan Naïve Bayes dengan metode *test percentage split 35%* yaitu dari data akan digunakan sebagai data training sementara 65% dari data akan digunakan sebagai data testing. Hasil aplikasi Weka dalam menganalisis data penggunaan internet dengan menggunakan metode *test percentage split 35%*. adalah sebagai berikut:



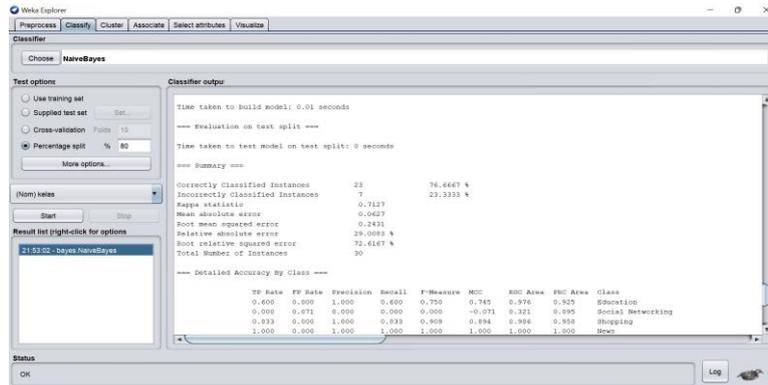
Gambar 3. Tampilan Form Hasil Analisis 35%

Analisis data Penggunaan internet menggunakan Naïve Bayes dengan metode *test percentage split 50%* maksudnya ialah 50% dari data akan digunakan sebagai data training sementara 50% dari data akan digunakan sebagai data testing. Hasilnya adalah sebagai berikut:



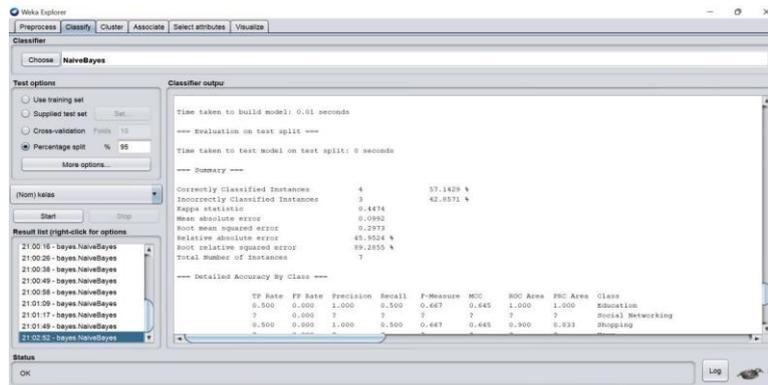
Gambar 4. Tampilan Form Hasil Analisis 50%

Analisis data Penggunaan internet menggunakan Naïve Bayes dengan metode *test percentage split 80%* maksudnya ialah 80% dari data akan digunakan sebagai data training sementara 20% dari data akan digunakan sebagai data testing. Hasilnya adalah sebagai berikut:



Gambar 5. Tampilan Form Hasil Analisis 80%

Analisis data Penggunaan internet menggunakan Naïve Bayes dengan metode *test percentage split* 95% maksudnya ialah 95% dari data akan digunakan sebagai data training sementara 5% dari data akan digunakan sebagai data testing. Hasil aplikasi Weka dalam menganalisis data penggunaan internet dengan menggunakan metode *test percentage split* 95%. adalah sebagai berikut:



Gambar 6. Tampilan Form Hasil Analisis 95%

Berdasarkan hasil yang sudah diuji menggunakan *Percentage Split* melalui aplikasi weka maka dapat disimpulkan melalui tabel dibawah ini.

Model <i>percentage split</i>	Data yang berhasil diklasifikasikan	Data yang tidak berhasil diklasifikasikan
20%	60.8333%	39.1667%
35%	72.1649%	27.8351%
50%	77.3333%	22.6667%
65%	71.1538%	28.8462%
80%	76.6667%	23.3333%
95%	57.1429%	42.8571%

Tabel 1 Hasil Kesimpulan Pengujian

Berdasarkan hasil diatas dapat disimpulkan Model *Percentage Split* 50% memiliki tingkat eror yang rendah dan Model *Percentage Split* 20% memiliki tingkat eror yang besar.

4. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan maka penulis mengambil beberapa kesimpulan, diantaranya adalah sebagai berikut :

- Metode *Naïve Bayes* memanfaatkan data *training* untuk menghasilkan probabilitas setiap kriteria untuk *class* yang berbeda, sehingga nilai-nilai probabilitas dari kriteria tersebut dapat dioptimalkan untuk proses klasifikasi yang dilakukan oleh metode *Naïve Bayes* itui sendiri.
- Berdasarkan hasil yang telah diuji menggunakan metode *naive bayes* mennggunakan metode test *Percentage Split*, dapat disimpulkan metode test *Percentage Split* 50% yang terbaik dikarenakan memiliki tingkat eror yang rendah dibandingkan lainnya.

DAFTAR PUSTAKA

- H. Muhamad, C. A. Prasajo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, "Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 180, 2017, doi: 10.25126/jtiik.201743251.
- D. Sartika and D. I. Sensuse, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 3, no. 2, pp. 151–161, 2017, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/78>.
- Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.
- D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- R. S. Sasmita, "Research & Learning in Primary Education Pemanfaatan Internet Sebagai Sumber Belajar," *J. Pendidik. Dan Konseling*, vol. 1, pp. 1–5, 2020.
- I. Amarini, "Pencegahan Dampak Negatif Perkembangan Teknologi Informasi Terhadap Pengguna Internet," *Kosmik Huk.*, vol. 18, no. 1, 2018, doi: 10.30595/kosmikhukum.v18i1.2340.
- D. P. Musdansi and R. Nazli, "Pengembangan Buku Ajar Statistika Berbasis Spss Sebagai Self Education Mahasiswa," *AdMathEdu J. Ilm. Pendidik. Mat. Ilmu Mat. dan Mat. Terap.*, vol. 8, no. 2, p. 147, 2018, doi: 10.12928/admathedu.v8i2.12342.
- Ainurrohmah, "Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka," *Prisma*, vol. 4, pp. 493–499, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>.
- D. Yanti, "Perencanaan Prototipe Sistem Informasi Pariwisata Berbasis Web Di Kabupaten Dairi Sumatera Utara," vol. 7, no. 2, pp. 44–68, 2018