Seminar Nasional Fakultas Ekonomi dan Sains Teknologi (SINERGITEK) *ISSN: 3074 - 1354*
Vol. 1 No. 1 (2023)
https://doi.org/

Metode Klasifikasi dalam Data *Mining* untuk Memprediksi Keterlambatan Pembayaran SPP

Petti Indrayati Sijabat[1*], Fitri Aini Nasution[2]

Sekolah Tinggi Manajemen dan Ilmu Komputer, Prodi Teknik Informatika

Abstrak

Keterlambatan pembayaran SPP siswa MTsS Al-Islamiyah Karang Anyar merupakan hal penting yang perlu diperhatikan, karena sangat berpengaruh terhadap keuangan sekolah. Metode Klasifikasi merupakan suatu metode yang terdapat pada Data Mining, digunakan untuk mengelompokkan data. Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari komputer. Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma, diantaranya Naïve Bayes, Support Vector Machine, Decission Tree, Fuzzy dan JST. Algoritma yang digunakan pada penelitian ini adalah algoritma C4.5. Hasil dari penelitian ini ialah berhasil membuat sistem prediksi berbasis website yang dapat digunakan untuk memprediksi pembayaran SPP pada MTsS Al-Islamiyah Karang Anyar. Dengan menggunakan metode klasifikasi yaitu Algoritma C4.5, telah diketahui pola data pada data training yang telah di input ke dalam sistem. Selanjutnya, sistem bekerja dengan mengklasifikasikan data baru sesuai input yang diberikan dengan membaca pola data tersebut.Aplikasi web prediksi keterlambatan pembayaran SPP pada MTsS Al-Islamiyah Karang Anyar menggunakan metode Klasifikasi (Algoritma C4.5) memiliki keluaran (output) berupa informasi penentuan label atau kelas tiap-tiap data siswa, dengan diketahuinya kelas data tersebut memudahkan pihak MTsS Al-Islamiyah Karang Anyar untuk memperkirakan / memprediksi siswa yang akan berpotensi tidak lancar melakukan pembayaran SPP.

Kata Kunci : Data Mining, C4.5, Klasifikasi, Prediksi

Abstrack

The delay in paying tuition fees for MTsS Al-Islamiyah Karang Anyar students is an important matter that needs attention, because it greatly affects school finances. The Classification Method is a method found in Data Mining, used to group data. Classification that is done manually is a classification that is done by humans without any assistance from a computer. While the classification is done with the help of technology, has several algorithms, including Naïve Bayes, Support Vector Machine, Decision Tree, Fuzzy and ANN. The algorithm used in this study is the C4.5 algorithm. The result of this research is to succeed in creating a website-based prediction system that can be used to predict tuition payments at MTsS Al-Islamiyah Karang Anyar. By using the classification method, namely Algorithm C4.5, it is known that the data patterns in the training data that have been input into the system. Furthermore, the system works by classifying new data according to the input provided by reading the data pattern. The web application predicts late payment of tuition fees at MTsS Al-Islamiyah Karang Anyar using the Classification method (Algorithm C4.5) has an output in the form of information on determining labels or the class of each student data, knowing the class of the data makes it easier for MTsS Al-Islamiyah Karang Anyar to estimate/predict students who will potentially not be able to pay tuition fees smoothly.

Keywords : Data Mining, C4.5, Classification, Prediction

1. Pendahuluan

Dengan seiring perkembangan zaman, teknologi di dunia saat ini juga ikut berkembang dan semakin maju, sehingga mesyarakat di seluruh dunia dituntut untuk mengikuti kemajuan teknologi atau biasa kita sebut dengan istilah IPTEK (Ilmu Pengetahuan dan Teknologi) yang ada saat ini bahkan kedepannya. Teknologi saat ini sudah mencakup ke berbagai bidang, terutama dalam bidang pendidikan. Pada setiap sekolah saat ini telah menerapkan penggunaan teknologi informasi untuk membantu para pegawai khususnya operator sekolah untuk mengolah data para siswa, termasuk data keuangan. Sebagai contoh penggunaannya ialah untuk memprediksi keterlambatan pembayaran SPP.

* Corresponding author.
E-mail address: petti.jabat@gmail.com (First Author)

SINERGITEK is licensed under an
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Keterlambatan pembayaran SPP siswa MTsS Al-Islamiyah Karang Anyar merupakan hal penting yang perlu diperhatikan, karena sangat berpengaruh terhadap keuangan sekolah. Keterlambatan pembayaran SPP biasanya dipengaruhi oleh faktor latar belakang ekonomi orangtua siswa yang kurang mampu. Adapun siswa yang sengaja menunda-nunda pembayaran padahal orang tuanya mampu, hal ini bisa dilihat dari data siswa tersebut. Berdasarkan keadaan di lapangan kebanyakan siswa yang sering menunggak berasal dari kelas 9, karena bisa dilihat ketika akan melaksanakan ujian banyak yang baru membayar disaat itu juga. Dari masalah tersebut, tentunya berdampak terhadap biaya pengadaan fasilitas sekolah serta honor tenaga pendidik. Maka, diperlukan sistem yang dapat memprediksi keterlambatan pembayaran SPP untuk mengetahui berapa banyak siswa yang berpotensi akan menunggak, serta memilah siswa yang kemungkinan sengaja menunda pembayaran berdasarkan data yang ada. Kemudian, pihak sekolah bisa mengambil tindakan jika diketahui siswa tersebut benar sengaja menunda pembayaran, seperti mempertemukan keluarga siswa dengan pihak sekolah sesuai dengan kebijakan lembaga pendidikan.

Lembaga pendidikan berhak untuk memberikan sanksi berupa ketidakikutsertaan ujian bagi siswa yang belum membayar SPP. Keterlambatan pembayaran SPP MTsS Al-Islamiyah Karang Anyar belum pernah diprediksi sebelumnya, namun sering terjadi penunggakan oleh beberapa siswa. Maka, diperlukan aplikasi yang dapat memprediksi tingkat keterlambatan pembayaran SPP dengan menerapkan metode Klasifikasi.

Metode Klasifikasi merupakan suatu metode yang terdapat pada Data Mining, digunakan untuk mengelompokkan data. Klasifikasi yang dilakukan secara manual adalah klasifikasi yang dilakukan oleh manusia tanpa adanya bantuan dari algortima cerdas komputer. Sedangkan klasifikasi yang dilakukan dengan bantuan teknologi, memiliki beberapa algoritma, diantaranya Naïve Bayes, Support Vector Machine, Decission Tree, Fuzzy dan Jaringan Saraf Tiruan[1]. Algoritma yang digunakan pada penelitian ini adalah algoritma C4.5. Metode klasifikasi algoritma C4.5 sebelumnya pernah digunakan untuk penelitian berjudul "Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi Decision Tree dan Support Verctor Machine"[2]. Dari penelitian tersebut telah diuji coba dengan software WEKA membuktikan bahwa algoritma C4.5 dan SMV mampu menyelesaikan kasus tersebut dengan nilai akurasi 70%, dengan hasil atribut Tanggal Penerimaan Gaji merupakan faktor utama penyebab keterlambatan konsumen dalam membayar tagihan listrik. Ada juga penelitian yang menggunakan metode algoritma ini berjudul "Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP)"[3].

Berdasarkan data penelitian tersebut telah diuji dengan perhitungan manual dan perangkat lunak RapidMiner, hasil accuracy mencapai 90,24% diperoleh dari data training dan data testing. Adapun penelitian lain yang menggunakan C4.5 dengan judul "Implementasi Algoritma C4.5 Untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python"[4]. Hasil dari penelitian tersebut yang diterapkan ke premrograman Python didapat nilai akurasi sebesar 73%. Dibandingkan dengan penelitian sebelumnya, penilitian saat ini berbeda studi kasus, atribut-atribut kriteria, dan implementasi program yang digunakan, maka output-nya sudah pasti berbeda. Hasil penelitian akan membuktikan penerapan Data Mining dengan menggunakan metode klasifikasi dapat membantu prediksi para siswa yang akan terlambat membayar SPP, sehingga pihak sekolah dapat memperkirakan kemungkinan yang akan terjadi, seperti mengetahui siswa yang benar belum mampu membayar dengan siswa yang sengaja tidak membayar.

Dengan beberapa pertimbangan sesuai dengan pemaparan diatas, maka diangkatlah judul penelitian ini yaitu Penerapan Data Mining Menggunakan Metode Klasifikasi Untuk Prediksi Keterlambatan Pembayaran SPP Siswa MTsS Al-Islamiyah Karang Anyar. Harapannya agar masalah seperti ini dapat teratasi dengan mudah dan cepat, serta memberikan bukti nyata bahwa penerapan teknologi di bidang pendidikan sangat bermanfaat.

2. Landasan Teori

Klasifikasi adalah teknik pengolahan data yang membagi suatu objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan[5]. Algoritma C4.5 adalah algoritma yang digunakan untuk membangun pohon keputusan, metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mewakili aturan dalam bahasa alami yang mudah dipahami[6].

Beberapa tahapan dari algoritma C4.5 adalah:

1. Menyiapkan data yang akan dikelompokkan ke dalam kelas-kelas tertentu.

2. Tentukan akar yang dibentuk oleh pohon keputusan. Akar diperoleh dari variabel yang dipilih dengan menghitung Gain untuk setiap atribut.
3. Setelah nilai Gain untuk masing-masing variabel diperoleh, maka dipilih atribut dengan nilai Gain terbesar.
4. Ulangi langkah-langkah tersebut hingga semua record memiliki kelas yang sama.

Sebuah pohon keputusan dimulai dari akar paling atas, jika satu set data uji diberikan. Misalkan kelas data y tidak diketahui, pohon keputusan ditelusuri dari akar ke node dan setiap nilai atribut sesuai data y diuji untuk melihat apakah mengikuti aturan pohon keputusan, selanjutnya pohon keputusan memprediksi kelas dari sebuah tuple y. Algoritma C4.5 dan pohon keputusan merupakan dua model yang tidak dapat dipisahkan, karena algoritma C4.5 diperlukan untuk membangun pohon keputusan.

Entropy adalah ukuran dari teori informasi yang dapat mengetahui karakteristik dari impuryt dan homogenity dari kumpulan data. Dari nilai Entropy tersebut kemudian dihitung nilai information gain masing-masing atribut. Penghitungan nilai Entropy digunakan rumus seperti dalam Persamaan [5].

$$Entropy(\boldsymbol{S}) = \sum_{i=0}^{n} - pi * log2(pi) \quad (1)$$

Rumus (1) merupakan rumus yang digunakan dalam perhitungan entropy yang digunakan untuk menentukan seberapa informatif atribut tersebut. Berikut keterangannya :

s : Himpunan kasus
n : Jumlah partisi
pi : Jumlah kasus pada partisi ke-i

Information Gain adalah informasi yang didapatkan dari perubahan entropy pada suatu kumpulan data, baik melalui observasi atau bisa juga disimpulkan dengan cara melakukan partisipasi terhadap suatu set data.

$$GAIN(S, A) = Entropy(S) - \sum_{i=1}^{n} - \frac{|Si|}{|S|} * Entropy(S) \quad (2)$$

Rumus (2) merupakan rumus yang digunakan dalam perhitungan information gain setelah melakukan perhitungan entropy. Berikut keterangannya :

s : Himpunan kasus
n : Jumlah partisi atribut A
|Si | : Jumlah kasus pada partisi ke-i
|S| : Jumlah kasus dalam s

Dengan mengetahui rumus-rumus diatas, data yang telah diperoleh dapat dimasukkan dan diproses dengan algoritma C4.5 untuk proses pembuatan *decision tree.*

3. Metode Penelitian

Kerangka kerja adalah struktural konseptual dasar yang digunakan untuk memecahkan atau memproses masalah yang kompleks. Kerangka ini merupakan langkah-langkah yang diambil dalam penyelesaian pembahasan. Adapun kerangka kerja penelitian ini dapat digambarkan pada Gambar 1.

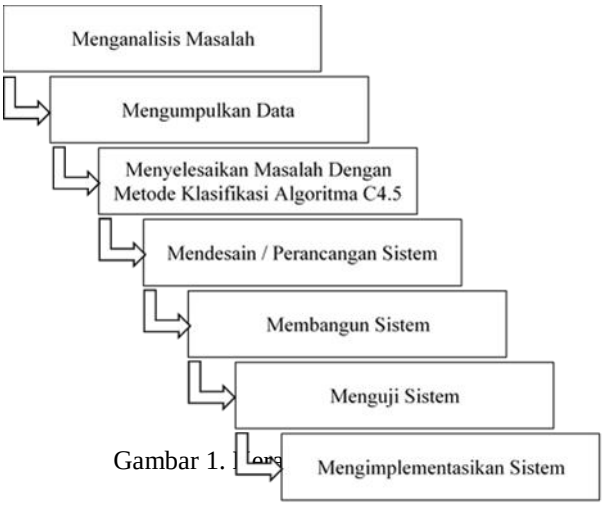
Berikut ini merupakan uraian kerangka kerja penelitian:

1. Menganalisis Masalah

Menganalisis masalah dalam penulisan penelitian ini adalah mengamati hal-hal yang menyebabkan timbulnya masalah, serta mencari solusi penyelesaian masalah tersebut.

2. Mengumpulkan Data

Dalam merancang sistem ini dibutuhkan teknik pengumpulan data untuk mendukung kegiatan penelitian ini, sebagai berikut:



Gambar 1. [illegible]

a. Wawancara (Interview), yaitu melakukan sesi tanya jawab langsung dengan pimpinan MTsS Al-Islamiyah Karang Anyar untuk mendapatkan informasi dari objek yang akan diteliti, khususnya untuk mendapatkan data yang bersifat internal.
b. Kuesioner/Angket, yaitu memberikan seperangkat pertanyaan atau pernyataan kepada responden untuk dijawab secara tertulis.

3. Menyelesaikan Masalah Dengan Metode Klasifikasi Algoritma C4.5

Dalam menyelesaikan masalah pada penelitian ini menggunakan metode klasifikasi yaitu algoritma Decision Tree (C.45). Dengan metode ini akan membantu mengatasi masalah tersebut, dan masuk ketahap penelitian selanjutnya.

4. Mendesain / Perancangan Sistem

Perancangan sistem merupakan proses perancangan untuk merancang sistem atau memperbaiki sistem yang sudah ada agar bekerja lebih baik, lebih efektif dan efisien. Proses desain disajikan dalam bentuk desain input, desain output, dan desain file.

5. Membangun Sistem

Setelah melakukan perancangan sistem yang akan dibuat, selanjutnya pada tahap ini akan dibangun sistem. Dalam membangun sistem diperlukan perencanaan yang baik seperti menganalisis kendala yang mungkin terjadi pada proses pembuatan sistem, mengumpulkan data, dan membuat rancangan atau desain tentang sitem yang dibangun untuk dapat diuji serta diimplementasikan.

6. Menguji Sistem

Pengujian sistem merupakan proses eksekusi untuk menemukan dan memperbaiki kesalahan pada sistem. Ini merupakan tahap penting dalam pengembangan sistem karena merupakan tahap di mana perlu dipastikan bahwa sistem bebas dari kesalahan.

7. Mengimplementasikan Sistem

Implementasi sistem adalah fase ketika sistem siap untuk dioperasikan. Setelah dipastikan semua tahapan telah terpenuhi, pada tahap ini dilakukan penerapan/digunakan secara langsung sistem tersebut.

4. Hasil dan Pembahasan

Adapun sampel data latih yang dibutuhkan untuk merancang sistem ini dapat dilihat pada Tabel 1. Sampel data ini diambil berdasarkan jumlah data latih yang akan digunakan pada sistem nantinya sebanyak 100 data.

Keterangan atribut:

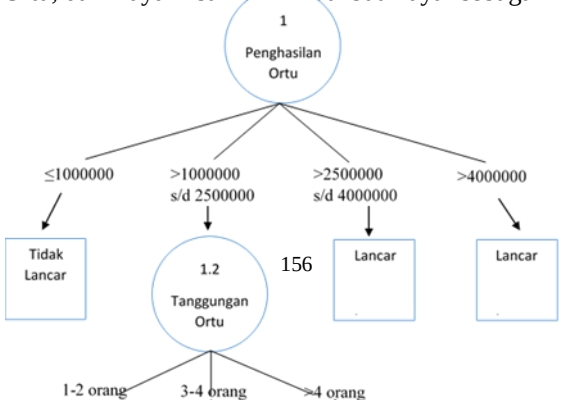
Tanggungan Orangtua = Tanggungan Ortu

Penghasilan Orangtua = Penghasilan Ortu

Tabel 1. Sampel Atribut Data Pelatihan

No	NISN	Nama	Penerima PIP/KIP	Tanggungan Ortu	Penghasilan Ortu	Daya Listrik	Bayar
1	0089934298	MHD LUTFI AL HAFIEZ	Tidak	3-4 orang	>4000000	900 VA	Lancar
2	0087704336	RITA PUSPITA SARI	Tidak	1-2 orang	>2500000 s/d 4000000	900 VA	Lancar
3	0058398442	RENDI SAPUTRA	Tidak	1-2 orang	>4000000	1300 VA	Lancar
4	0072660118	ANDIKA ABDEE NEGARA	Tidak	>4 orang	>1000000 s/d 2500000	900 VA	Tidak Lancar
5	0088518059	SABILA SYAROFAH	Tidak	3-4 orang	>1000000 s/d 2500000	450 VA	Tidak Lancar
6	0082358548	KHOSIATIN WAHYUNI	Ya	3-4 orang	>1000000 s/d 2500000	900 VA	Lancar
7	0071158665	ALFI NUR HIDAYAH	Tidak	3-4 orang	>1000000 s/d 2500000	>=2200 VA	Lancar
8	0099614904	MICHAEL ZEBUA	Tidak	3-4 orang	<=1000000	450 VA	Tidak Lancar
9	0089212972	KESKA ARITA PUTRI	Tidak	1-2 orang	>1000000 s/d 2500000	450 VA	Lancar
10	0076277309	RAHMAYANI	Ya	3-4 orang	>2500000 s/d 4000000	450 VA	Lancar
11	0076084715	REIHAN PRATAMA	Tidak	3-4 orang	<=1000000	900 VA	Tidak Lancar
12	0086549910	BAGUS TRIADI	Tidak	1-2 orang	>2500000 s/d 4000000	900 VA	Lancar
13	0085941985	FITRI WULANDARI	Tidak	1-2 orang	<= 1000000	900 VA	Tidak Lancar
14	0089787762	MUHAMMAD AFIQ	Tidak	3-4 orang	>4000000	1300 VA	Lancar
15	0088612805	AULIA SAFITRI	Ya	1-2 orang	<=1000000	900 VA	Tidak Lancar

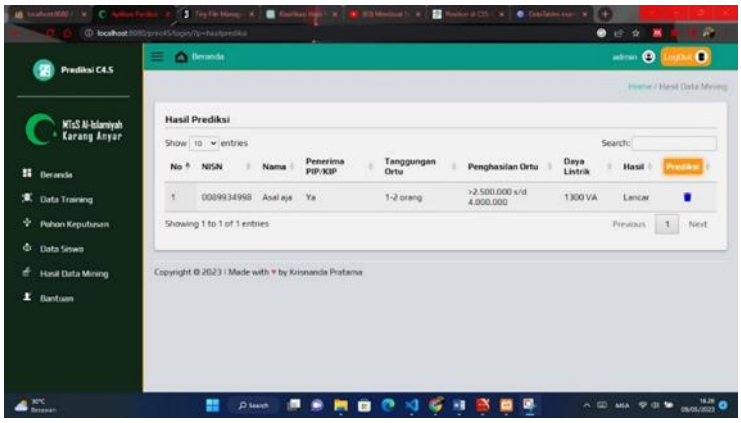
Berdasarkan Tabel 1, ada 4 atribut kriteria yang akan digunakan sebagai parameter yaitu Penerima PIP/KIP, Tanggungan Ortu, Penghasilan Ortu, dan Daya Listrik, serta atribut Bayar sebagai Label Kelas / Target.



Gambar 2. Hasil Pohon Keputusan

Berdasarkan sampel data yang telah diproses dengan menggunakan Algoritma C4.5, sudah didapat pohon keputusannya pada gambar di atas. Kesimpulannya, dari pola pohon keputusan di atas telah diketahui alur penentuan kelas jika ada data baru yang dimasukkan (input). Dalam hal ini telah diketahui bahwa atribut Penghasilan Ortu sebagai sebagai node akar. Apabila penghasilan ortu bernilai ⩽1000000 maka diklasifikasikan menjadi "Tidak Lancar", jika bernilai >2500000 s/d 4000000 maka diklasifikasikan menjadi "Lancar", jika bernilai >4000000 maka diklasifikasikan menjadi "Lancar", sedangkan untuk nilai >1000000 s/d 2500000 belum dapat diklasifikasikan karena ada beberapa pertimbangan lagi yaitu melihat "Tanggungan Ortu" sebagai node cabang. Apabila tanggungan ortu bernilai 1-2 orang maka diklasifikasikan menjadi "Lancar", begitu pula seterusnya.

Hasil mining akan ditampilkan pada halaman ini, dimana terdapat tombol prediksi untuk menambahkan data baru yang belum diketahui kelasnya, dalam kata lain akan diprediksi. Antarmuka halaman hasil data mining dapat dilihat pada gambar di bawah ini.



Gambar 3. Antarmuka Halaman Hasil Data Mini

5. Kesimpulan

Kesimpulan yang dapat diambil dari penelitian yang telah dilakukan yaitu sebagai berikut :

1. Aplikasi web prediksi keterlambatan pembayaran SPP pada MTsS Al-Islamiyah Karang Anyar menggunakan metode Klasifikasi (Algoritma C4.5) memiliki keluaran (output) berupa informasi penentuan label atau kelas tiap-tiap data siswa, dengan diketahuinya kelas data tersebut memudahkan pihak MTsS Al-Islamiyah Karang Anyar untuk memperkirakan / memprediksi siswa yang akan berpotensi tidak lancar melakukan pembayaran SPP. Jadi, ketika diketahui ada siswa yang banyak melakukan penunggakkan pembayaran SPP, dalam kondisi ini sesuai data yang telah diprediksi bahwa siswa yang bersangkutan termasuk kategori yang akan lancar melakukan pembayaran, maka siswa tersebut perlu dipertanyakan. Jika diketahui siswa yang bersangkutan sengaja tidak membayar, mata pihak sekolah berhak melakukan tindakkan seperti mempertemukan keluarga siswa dengan pihak sekolah.
2. Telah berhasil merancang dan menerapkan sistem prediksi menggunakan metode Klasifikasi (Algoritma C4.5) berbasis aplikasi web yang menghasilkan pohon keputusan, dimana dari bentuk pohon keputusan itu bisa didapatkan informasi pola data untuk menentukan kelas data baru sesuai dengan atributnya.

6. Daftar Pustaka

[1] A. P. Wibawa, M. G. A. Purnama, M. F. Akbar, and F. A. Dwiyanto, “Metode-metode Klasifikasi,” Pros. Semin. Ilmu Komput. dan Teknol. Inf., vol. 3, no. 1, p. 134, 2018.

[2] D. N. Batubara et al., “Analisis Prediksi Keterlambatan Pembayaran Listrik Menggunakan Komparasi Metode Klasifikasi Decision Tree dan Support Vector Machine,” J. Ris. Komput., vol. 9, no. 1, pp. 102–108, 2022, [Online]. Available: http://www.stmik-budidarma.ac.id/ejurnal/index.php/jurikom/article/view/3833

[3] M. Yunus, H. Ramadhan, D. R. Aji, and A. Yulianto, “Penerapan Metode Data Mining C4.5 Untuk Pemilihan Penerima Kartu Indonesia Pintar (KIP),” Paradig. - J. Komput. dan Inform., vol. 23, no. 2, 2021, doi: 10.31294/p.v23i2.11395.

[4] V. S. Ginting, K. Kusrini, and E. Taufiq, “Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python,” Inspir. J. Teknol. Inf. dan Komun., vol. 10, no. 1, 2020, doi: 10.35585/inspir.v10i1.2535.

[5] M. F. (Muhammad) Arifin and D. (Devi) Fitrianah, “Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada,” InComTech J. Telekomun. dan Komput., vol. 8, no. 2, pp. 87–102, 2018, [Online]. Available: https://www.neliti.com/publications/289475/

[6] M. A. Senubekti and L. A. P. Dewi, “Prinsip Klasifikasi Dan Data Mining Dengan Algoritma C4. 5,” Nuansa Inform., vol. 16, no. 2, pp. 87–93, Jul. 2022, doi: 10.25134/NUANSA.