

PENERAPAN METODE *DECISION TREE* C4.5 DAN *NAIVE BAYES* PADA KLASIFIKASI KARAKTERISTIK KEPRIBADIAN MANUSIA

Muhammad Hendry^{1*}, Yan Mitha Djaksana¹

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Jl. Raya Puspipetek No. 46, Kel. Buaran, Kec. Serpong, Kota Tangerang Selatan, Banten 15310, Indonesia

Email: ^{1*}hendry48net7@gmail.com, ²dosen01994@unpam.ac.id

(* : coressponding author)

Abstrak– Karakter pribadi seorang karyawan mempengaruhi sikap dan perilaku seseorang, termasuk sikap dan perilaku kerjanya. Dengan kata lain, kepribadian memiliki pengaruh terhadap kinerja karyawan, sehingga masuk akal bagi seorang atasan atau manajemen untuk terlebih dahulu memahami kondisi atau tipe kepribadian karyawan ketika berusaha meningkatkan kinerja karyawan. Di PT. Visi Insan Pratama sedang melakukan penambahan tim dan pegawai dengan melakukan wawancara dan tes tertulis, namun kedua cara tersebut masih belum cukup untuk mengenal kepribadian calon pegawai secara spesifik, sehingga masih banyak pertimbangan untuk menyeleksi calon pegawai baru. Algoritma C4.5 merupakan salah satu algoritma pembelajaran mesin. Dengan menggunakan algoritma ini, sebuah mesin (komputer) akan diberikan satu set data untuk dipelajari, yang disebut learning dataset. Algoritma Naive Bayes merupakan salah satu pengklasifikasi statistik yang dapat memprediksi probabilitas data yang dimiliki dengan menampilkan kelas dari kelas tertentu berdasarkan kemungkinan probabilitas. Sistem ini dapat menghasilkan informasi karakteristik kepribadian dengan menggunakan metode Decision Tree C4.5 dan Naive Bayes.

Kata Kunci: Kepribadian, Algoritma, Pohon Keputusan, C4.5, Naive Bayes

Abstract– *An employee's personal character influences a person's attitude and behavior, including his work attitude and behavior. In other words, personality has an influence on employee performance, so it makes sense for a boss or management to first understand the condition or type of employee personality when trying to improve employee performance. At PT. Insan Pratama's vision is to add to the team and employees by conducting interviews and written tests, but these two methods are still not enough to know the personality of prospective employees specifically, so there are still many considerations in selecting new prospective employees. The C4.5 algorithm is one of the machine learning algorithms. By using this algorithm, a machine (computer) will be given a set of data to study, which is called a learning dataset. The Naive Bayes algorithm is one of the statistical classifiers that can predict the probability of data being owned by displaying classes from certain classes based on possible probabilities. This system can produce information on personality characteristics using the Decision Tree C4.5 and Naive Bayes methods.*

Keywords: Personality, Algorithm, Decision Tree, C4.5, Naive Bayes

1. PENDAHULUAN

Karakter pribadi seorang karyawan mempengaruhi sikap dan perilaku seseorang, termasuk sikap dan perilaku kerjanya. Dengan kata lain, kepribadian memiliki pengaruh terhadap kinerja karyawan, sehingga masuk akal bagi seorang atasan atau manajemen untuk terlebih dahulu memahami kondisi atau tipe kepribadian karyawan ketika berusaha meningkatkan kinerja karyawan.

Pemahaman terhadap kepribadian karyawan menentukan strategi pengembangan karyawannya. Kepribadian individu dapat dilihat dari beberapa konsep penting seperti *locus of control*, *self efficacy*, dan *self concept*. Ketiganya terbentuk melalui proses pembelajaran dan dipengaruhi pengalaman dan lingkungan.

Di PT. Visi Insan Pratama sedang melakukan penambahan tim dan pegawai dengan melakukan wawancara dan tes tertulis, namun kedua cara tersebut masih belum cukup untuk mengenal kepribadian calon pegawai secara spesifik, sehingga masih banyak pertimbangan untuk menseleksi calon pegawai baru.

Manusia mempunyai karakter yang berbeda-beda dan unik. Karakter atau kepribadian manusia bisa dipelajari, dan manusia kadang memiliki kesamaan karakter antara satu dengan yang lainnya. Kepribadian manusia telah dikaji dan dirangkum menjadi empat jenis. Keempatnya masuk

dalam teori proto-psikologis, di mana teori itu dibagi lagi menjadi empat tipe kepribadian mendasar, yaitu sanguinis, melankolis, koleris dan plegmatis.

Pohon keputusan adalah metode klasifikasi yang kuat dan terkenal. Metode pohon keputusan mengubah fakta besar menjadi pohon keputusan yang mewakili aturan, membuat aturan mudah dipahami manusia. Pohon keputusan juga digunakan untuk mengeksplorasi data dan menemukan hubungan tersembunyi antara banyak variabel input dan variabel target. Keputusan dapat dibangun dengan menerapkan salah satu algoritma pohon keputusan untuk memodelkan sekumpulan data yang tidak terklasifikasi. Konsep pohon keputusan adalah mengubah data menjadi pohon keputusan dan aturan keputusan.

Algoritma C4.5 merupakan salah satu algoritma pembelajaran mesin. Dengan menggunakan algoritma ini, sebuah mesin (komputer) akan diberikan satu set data untuk dipelajari, yang disebut learning dataset. Hasil pembelajaran selanjutnya digunakan untuk mengolah data baru yang disebut dataset uji. Karena algoritma C4.5 digunakan untuk klasifikasi, maka hasil pengolahan dataset uji berupa pengelompokan data ke dalam kelas-kelasnya.

Algoritma *Naive Bayes* merupakan salah satu pengklasifikasi statistik yang dapat memprediksi probabilitas data yang dimiliki oleh keanggotaan kelas dari kelas tertentu berdasarkan perhitungan probabilitas.

2. METODOLOGI PENELITIAN

2.1 Metode Pengumpulan Data

a. Studi Pustaka

Studi pustaka dilakukan dengan cara membaca buku yang berkaitan dengan penelitian Tugas Akhir, lalu mempelajarinya sehingga penulis dapat mengetahui mengenai langkah dalam membuat Tugas Akhir yang baik dan benar.

b. Observasi

Observasi dilakukan dengan cara mengamati langsung ketempat tujuan, sehingga penulis akan mendapatkan informasi dan data yang dibutuhkan. Lokasi observasi yaitu SMK Era Pembangunan.

c. Wawancara

Wawancara dilakukan dengan cara datang langsung ke sekolah dan bertanya kepada kesiswaan yaitu Bapak Pandu Yuda Wardana S.M untuk bertanya tentang hal yang berkaitan dengan judul Tugas Akhir.

2.2 Metode Analisis

2.2.1 *Decision Tree C4.5*

Alasan mengapa penulis menggunakan *extreme programming* yaitu metode ini pada dasarnya diperuntukkan kepada pembangunan perangkat lunak yang kebutuhannya tidak tetap atau selalu berubah-ubah, artinya *extreme programming* merupakan metode yang bersifat responsive terhadap perubahan, sehingga perubahan kebutuhan akan selalu diterima oleh developer, tidak seperti pada model pengembangan perangkat lunak terstruktur lainnya. Selain berfokus pada *coding*, *extreme programming* juga mencakup seluruh aspek pengembangan perangkat lunak seperti perencanaan, desain, dan lain sebagainya.

2.2.2 *Naive Bayes*

Naive Bayes merupakan sebuah metode penggolongan berdasarkan probabilitas sederhana dan dirancang untuk dipergunakan dengan asumsi bahwa antar satu kelas dengan kelas yang lain tidak saling tergantung (independen). Pada klasifikasi *Naive Bayes*, proses pembelajaran lebih ditekankan pada mengestimasi probabilitas. Keuntungan dari pendekatan ini yaitu pengklasifikasian akan mendapatkan nilai *error* yang lebih kecil ketika *data set* berjumlah besar (Berry,2006).

3. ANALISA DAN PEMBAHASAN

3.1 Analisa

Analisa sistem merupakan tahapan yang sangat penting dalam proses pembuatan sistem informasi, karena dalam tahapan ini kesatuan Sistem Informasi akan diuraikan menjadi tiap-tiap bagian komponen sesuai dengan kebutuhannya. Dengan uraian tiap-tiap bagian tersebut akan mempermudah dalam proses evaluasi kebutuhan sistem, permasalahan serta hambatan yang akan terjadi sehingga dapat dilakukan perbaikan-perbaikan untuk menyempurnakan Sistem Informasi yang akan dibuat.

3.1.1 Analisa *Decision Tree* C4.5

Tabel 1. Pohon Keputusan

ID	Aturan
1	IF (jawaban_a<=10) AND (jawaban_c<=10) AND (jawaban_b<=10) THEN Label = Plegmatis
2	IF (jawaban_a<=10) AND (jawaban_c<=10) AND (jawaban_b>10) AND (usia='23') THEN Label = Koleris
3	IF (jawaban_a<=10) AND (jawaban_c<=10) AND (jawaban_b>10) AND (usia='25' OR usia='24') AND (jawaban_d<=10) THEN Label = Koleris
4	IF (jawaban_a<=10) AND (jawaban_c<=10) AND (jawaban_b>10) AND (usia='25' OR usia='24') AND (jawaban_d>10) AND (usia='24') AND (jawaban_a<=5) THEN Label = Koleris
5	IF (jawaban_a<=10) AND (jawaban_c<=10) AND (jawaban_b>10) AND (usia='25' OR usia='24') AND (jawaban_d>10) AND (usia='24') AND (jawaban_a>5) THEN Label = Plegmatis
6	IF (jawaban_a<=10) AND (jawaban_c<=10) AND (jawaban_b>10) AND (usia='25' OR usia='24') AND (jawaban_d>10) AND (usia='25') THEN Label = Koleris
7	IF (jawaban_a<=10) AND (jawaban_c>10) AND (jawaban_d<=10) THEN Label = Melankolis
8	IF (jawaban_a<=10) AND (jawaban_c>10) AND (jawaban_d>10) AND (jawaban_c<=15) AND (jenis_kelamin='L') AND (usia='23') THEN Label = Melankolis
9	IF (jawaban_a<=10) AND (jawaban_c>10) AND (jawaban_d>10) AND (jawaban_c<=15) AND (jenis_kelamin='L') AND (usia='25' OR usia='24') AND (jawaban_b<=5) THEN Label = Melankolis
10	IF (jawaban_a<=10) AND (jawaban_c>10) AND (jawaban_d>10) AND (jawaban_c<=15) AND (jenis_kelamin='L') AND (usia='25' OR usia='24') AND (jawaban_b>5) THEN Label = Plegmatis
11	IF (jawaban_a<=10) AND (jawaban_c>10) AND (jawaban_d>10) AND (jawaban_c<=15) AND (jenis_kelamin='P') THEN Label = Plegmatis
12	IF (jawaban_a<=10) AND (jawaban_c>10) AND (jawaban_d>10) AND (jawaban_c>15) THEN Label = Melankolis
13	IF (jawaban_a>10) AND (usia='23') AND (jawaban_c<=10) AND (jawaban_d<=15) AND (pengalaman='Ya') THEN Label = Sanguin
14	IF (jawaban_a>10) AND (usia='23') AND (jawaban_c<=10) AND (jawaban_d<=15) AND (pengalaman='Tidak') AND (jawaban_d<=10) THEN Label = Sanguin

15	IF (jawaban_a>10) AND (usia='23') AND (jawaban_c<=10) AND (jawaban_d<=15) AND (pengalaman='Tidak') AND (jawaban_d>10) AND (jenis_kelamin='L') THEN Label = Plegmatis
16	IF (jawaban_a>10) AND (usia='23') AND (jawaban_c<=10) AND (jawaban_d<=15) AND (pengalaman='Tidak') AND (jawaban_d>10) AND (jenis_kelamin='P') AND (jawaban_b<=5) THEN Label = Plegmatis
17	IF (jawaban_a>10) AND (usia='23') AND (jawaban_c<=10) AND (jawaban_d<=15) AND (pengalaman='Tidak') AND (jawaban_d>10) AND (jenis_kelamin='P') AND (jawaban_b>5) THEN Label = Sanguin
18	IF (jawaban_a>10) AND (usia='23') AND (jawaban_c<=10) AND (jawaban_d>15) THEN Label = Plegmatis
19	IF (jawaban_a>10) AND (usia='23') AND (jawaban_c>10) THEN Label = Melankolis
20	IF (jawaban_a>10) AND (usia='24' OR usia='25') AND (jawaban_b<=10) AND (jawaban_c<=10) THEN Label = Sanguin
21	IF (jawaban_a>10) AND (usia='24' OR usia='25') AND (jawaban_b<=10) AND (jawaban_c>10) AND (usia='24') THEN Label = Sanguin
22	IF (jawaban_a>10) AND (usia='24' OR usia='25') AND (jawaban_b<=10) AND (jawaban_c>10) AND (usia='25') THEN Label = Sanguin
23	IF (jawaban_a>10) AND (usia='24' OR usia='25') AND (jawaban_b>10) AND (jawaban_a<=15) AND (jawaban_d<=5) THEN Label = Koleris
24	IF (jawaban_a>10) AND (usia='24' OR usia='25') AND (jawaban_b>10) AND (jawaban_a<=15) AND (jawaban_d>5) AND (jenis_kelamin='L') THEN Label = Sanguin
25	IF (jawaban_a>10) AND (usia='24' OR usia='25') AND (jawaban_b>10) AND (jawaban_a<=15) AND (jawaban_d>5) AND (jenis_kelamin='P') THEN Label = Sanguin
26	IF (jawaban_a>10) AND (usia='24' OR usia='25') AND (jawaban_b>10) AND (jawaban_a>15) THEN Label = Sanguin

3.1.2 Analisa Naïve Bayes

Data yang sudah melalui tahap preprocessing maka akan dijadikan data latih untuk mengklasifikasikan data uji menggunakan metode Naive Bayes. Dalam tahap ini akan dilakukan proses utama yaitu menghitung nilai probabilitas berdasarkan data latih. Selanjutnya menghitung nilai probabilitas tiap fitur berdasarkan data latih, menghitung nilai probabilitas akhir, dan selanjutnya data uji diklasifikasikan pada kelas dengan nilai probabilitas akhir terbesar.

1. Perhitungan Sample Data Uji Ke 1

Jenis kelamin: L Usia: 23 pengalaman: tidak Jawaban A: 12 Jawaban B: 1 Jawaban C: 12 Jawaban D: 15

Tabel 2. Uji 1

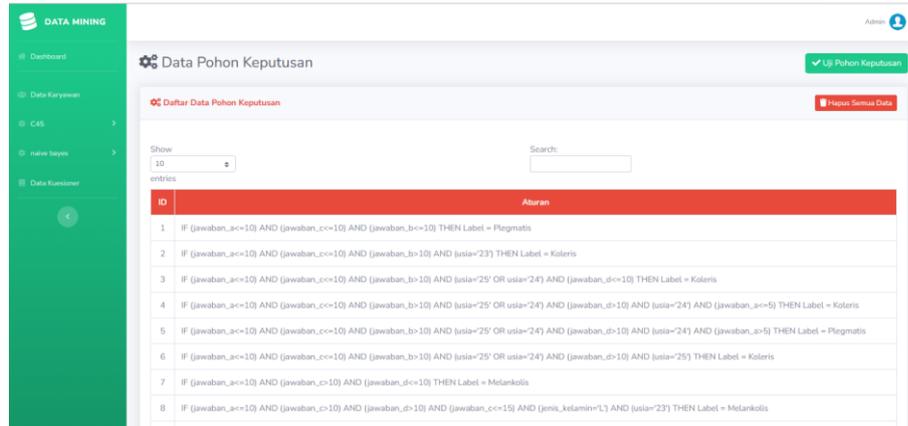
Jenis Kelamin:	Sanguin	Koleris	Melankolis	Plegmatis
Laki-laki	0.6333	0.6667	0.8333	0.56
Perempuan	0.3667	0.3333	0.1667	0.44

4. IMPLEMENTASI

4.1 Implementasi

4.1.1 Implementasi *Decision Tree* C4.5

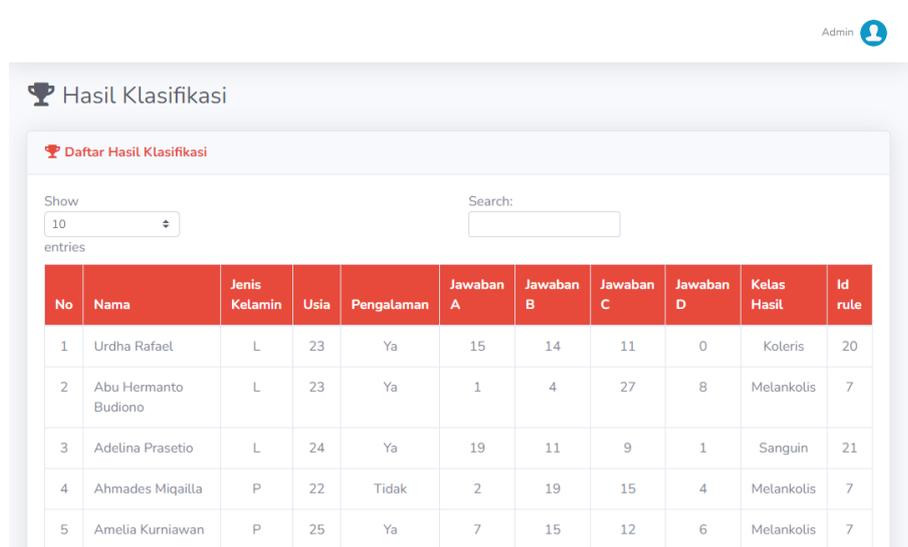
Pada Prototype output ini akan terlihat gambaran mengenai hasil masukan data yang telah dilakukan oleh pengguna pada setiap prosesnya dalam menggunakan sistem yang dibuat. Laporan merupakan berkas hasil dari proses yang telah dilakukan pada sistem. Adapun bentuknya dapat dilihat pada gambar.



ID	Aturan
1	IF (jawaban_a=<10) AND (jawaban_c=<10) AND (jawaban_b=<10) THEN Label = Plegmatik
2	IF (jawaban_a=<10) AND (jawaban_c=<10) AND (jawaban_b=10) AND (usia=23) THEN Label = Koleris
3	IF (jawaban_a=<10) AND (jawaban_c=<10) AND (jawaban_b=10) AND (usia=25' OR usia=24) AND (jawaban_d=<10) THEN Label = Koleris
4	IF (jawaban_a=<10) AND (jawaban_c=<10) AND (jawaban_b=10) AND (usia=25' OR usia=24) AND (jawaban_d=10) AND (usia=24) AND (jawaban_a=5) THEN Label = Koleris
5	IF (jawaban_a=<10) AND (jawaban_c=<10) AND (jawaban_b=10) AND (usia=25' OR usia=24) AND (jawaban_d=10) AND (usia=24) AND (jawaban_a=5) THEN Label = Plegmatik
6	IF (jawaban_a=<10) AND (jawaban_c=<10) AND (jawaban_b=10) AND (usia=25' OR usia=24) AND (jawaban_d=10) AND (usia=25) THEN Label = Koleris
7	IF (jawaban_a=<10) AND (jawaban_c>10) AND (jawaban_d=<10) THEN Label = Melankolis
8	IF (jawaban_a=<10) AND (jawaban_c>10) AND (jawaban_d=10) AND (jawaban_cc=15) AND (jenis_kelamin=L) AND (usia=23) THEN Label = Melankolis

Gambar 1. Tampilan Hasil Pohon Keputusan

Pada Prototype output ini akan terlihat gambaran mengenai hasil masukan data yang telah dilakukan oleh pengguna pada setiap prosesnya dalam menggunakan sistem yang dibuat. Laporan merupakan berkas hasil dari proses yang telah dilakukan pada sistem. Adapun bentuknya dapat dilihat pada gambar



No	Nama	Jenis Kelamin	Usia	Pengalaman	Jawaban A	Jawaban B	Jawaban C	Jawaban D	Kelas Hasil	Id rule
1	Urdha Rafael	L	23	Ya	15	14	11	0	Koleris	20
2	Abu Hermanto Budiono	L	23	Ya	1	4	27	8	Melankolis	7
3	Adelina Prasetyo	L	24	Ya	19	11	9	1	Sanguin	21
4	Ahmad M. Miqailla	P	22	Tidak	2	19	15	4	Melankolis	7
5	Amelia Kurniawan	P	25	Ya	7	15	12	6	Melankolis	7

Gambar 2. Tampilan Hasil Klasifikasi

4.1.2 Implementasi *Naïve Bayes*

Pada Prototype output ini akan terlihat gambaran mengenai hasil masukan data yang telah dilakukan oleh pengguna pada setiap prosesnya dalam menggunakan sistem yang dibuat. Laporan merupakan berkas hasil dari proses yang telah dilakukan pada sistem. Adapun bentuknya dapat dilihat pada gambar

Hasil:											
No	Nama	Jenis Kelamin	Usia	pengalaman	Jawaban A	Jawaban B	Jawaban C	Jawaban D	Kelas Asli	Kelas Hasil	
1	Abdul Majid	L	23	tidak	12	1	12	15	Plegmatis	Plegmatis	Benar
2	Abdul Rhaurf	L	24	ya	8	4	15	13	Melankolis	Melankolis	Benar
3	Ade Wahyu	L	24	ya	9	4	9	18	Plegmatis	Plegmatis	Benar
4	Adi Sugianto	L	24	ya	17	4	7	12	Sanguin	Sanguin	Benar
5	Aditya	L	24	ya	10	10	14	6	Melankolis	Melankolis	Benar
6	Aditya Herwian	L	24	ya	9	11	7	13	Plegmatis	Sanguin	Salah
7	Adnan Maulana	L	24	ya	9	4	10	17	Plegmatis	Plegmatis	Benar
8	Agus Hariyandi	L	24	ya	7	11	10	12	Plegmatis	Sanguin	Salah
9	Akbar	L	24	ya	14	6	11	9	Sanguin	Sanguin	Benar
10	Ali	L	23	tidak	7	6	15	12	Melankolis	Plegmatis	Salah
11	Angga Septian	L	25	ya	17	11	6	6	Sanguin	Sanguin	Benar

Gambar 3. Tampilan Hasil Uji Akurasi

Jumlah prediksi: 79

Jumlah tepat: 58

Jumlah tidak tepat: 21

AKURASI = 73.42 %

LAJU ERROR = 26.58 %

5. KESIMPULAN

Hasil yang didapatkan dari penelitian ini dapat disimpulkan sebagai berikut:

- Sistem ini dapat menghasilkan informasi karakteristik kepribadian dengan menggunakan metode *Decision Tree C4.5* dan *Naïve Bayes*.
- Komposisi data latih akan mempengaruhi tingkat akurasi yang didapatkan. Dimana setiap proses percobaan yang telah dilakukan akan mendapatkan akurasi yang berbeda-beda dan membentuk pohon keputusan yang terbaik.
- Data Latih yang digunakan dalam proses klasifikasi mempengaruhi hasil pengujian. Semakin banyak data latih yang digunakan dan semakin bervariasi data latih, hasil akurasi yang diperoleh akan semakin baik.
- Dari percobaan satu data yang dicoba beberapa kali dengan posisi pilihan yang berbeda tapi mendapatkan jumlah jawaban yang sama akan mendapatkan hasil yang sama.
- Pengujian pada sistem yang dilakukan dengan menginputkan data pegawai dan nama user yang berbeda namun jumlah nilai pada hasil jawaban kuesioner sama meskipun pengisian kuesioner dilakukan secara acak, hasil dari sistem menunjukkan klasifikasi pada kelas kepribadian yang sama.

REFERENCES

- Ginjar, T. (2014). Rahasia Membangun Website Toko Online Berpenghasilan Jutaan Rupiah. Iffahmedia.
- Heriyanto, Y. (2018). Perancangan Sistem Informasi Rental Mobil Berbasis WEB Pada PT.APM RENT CAR. 2(2), 64–77.
- Ikna, A. (2017). PERANCANGAN DAN PEMBUATAN APLIKASI E- COMMERCE PADA TOKO IKNA CHOCO CAKE.



- Ilhamsyah., & Shantika, M. (2015). , Rancang Bangun Aplikasi Sistem Informasi Pajak Bumi Bangunan.
- Mulyani, S. (2016). Metode Analisis dan Perancangan Sistem, ISBN: 978-979-19906-2-2: Vol. Edisi Ke-2. Abdi SisteMatika.
- Nuraizah, A. S. (2021). Rancang Bangun E- Learning Berbasis Website Pada SD Islam Nur Hidayah Batam.
- Porwati, P. (2020). Implementasi Kualitas Pelayanan Jasa Terhadap Kepuasan Pasien Pada Klinik As-Sakinah Tamansari Banyuwangi.
- Rosa, & Shalahuddin. (2018). Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek. In Informatika Bandung. Informatika.
- Sulistiyowati, Y. (2020). PERAN KEPEMIMPINAN, TANGGUNG JAWAB DAN KOMPENSASI TERHADAP LOYALITAS KARYAWAN DI RUMAH SAKIT UMUM 'AISYIYAH PONOROGO.
- Suparno. (2015). Rancangan Basis Data.
- Tabrani, M. (2014). Implementasi Sistem Informasi Reservasi Penginapan Pada Argowisata Gunung Mas Cisarua Bogor (Vol. 2, Issue 1). Bianglala Informatika.
- Widjajanto, N. (2012). *Sistem Informasi Akuntansi*. In Jakarta: Salemba Empat (Issue 33). PT Gelora Aksara Pratama.
- Yuni, S. (2015). Analisis dan Perancangan UML (Unefied Modelling Language) Generated VB.6. *Jurnal Teknik ITS*, 6(1), 1–6.