

Implementasi Algoritma Naive Bayes dan Algoritma C4.5 dalam Klasifikasi Kelayakan Bantuan UMKM

Aldy Novriandy*, Winarsih

Fakultas Teknologi Komunikasi dan Informatika, Program Studi Sistem Informasi, Universitas Nasional, Jakarta, Indonesia

Email: ^{1,*}aldynoriandy2019@student.unas.ac.id, ²Winarsih@civitas.unas.ac.id

Email Penulis Korespondensi: aldynoriandy2019@student.unas.ac.id

Abstrak—Perkembangan perekonomian Indonesia melambat akibat meningkatnya inflasi. Pengusaha sering menghadapi beberapa masalah dalam memulai suatu kegiatan kewirausahaan, termasuk: modal, perizinan, akuntansi panduan, promosi, barang, harga, sumber daya manusia, periklanan dan kegiatan lain yang sering menghambat proses bisnis. Pengelolaan Keuangan untuk Pemberdayaan UMKM adalah salah satu proses bisnis yang di jalankan oleh Kementerian Keuangan dalam meningkatkan kinerja UMKM melalui pemberian modal usaha. Namun dalam praktiknya, terdapat kendala dalam proses pengujian klasifikasi kelayakan terhadap kegiatan pemberian bantuan yang dilakukan, karena masih menggunakan cara manual yang tidak efisien. Agar pengujian yang dilakukan mampu lebih efektif dan praktis maka diperlukan perangkat lunak yang terintegrasi berbasis website. Untuk mencapai hal tersebut, penelitian ini dilakukan dengan bertujuan menghasilkan informasi terkait status kelayakan UMKM di seluruh wilayah Indonesia yang layak atau tidak layak menerima bantuan keuangan dari pemerintah. Agar informasi status kelayakan UMKM yang diperoleh dapat berguna bagi kantor wilayah di seluruh Indonesia, maka algoritma yang digunakan dalam penelitian ini adalah algoritma Naive Bayes dan algoritma C4.5. Hasil penelitian menunjukkan bahwa kedua algoritma tersebut dapat diterapkan dengan baik dalam penentuan kelayakan bantuan UMKM. Akurasi algoritma C4.5 adalah 90% sedangkan akurasi metode Naive Bayes adalah 70%. Algoritma C4.5 berkinerja sedikit lebih baik daripada Naive Bayes dalam pengaturan klasifikasi ini. Temuan akurasi dalam penelitian ini dapat dibandingkan dengan penelitian sebelumnya yang sudah dilakukan menggunakan algoritma yang sama atau dengan set data yang serupa. Hal ini membantu menentukan apakah metodologi penelitian ini menghasilkan tingkat akurasi yang lebih tinggi daripada penelitian sebelumnya.

Kata Kunci: UMKM; Website; Klasifikasi; Naive Bayes; C4.5

Abstract—Indonesia's economic development has slowed down due to rising inflation. Entrepreneurs often face several problems in starting an entrepreneurial activity, including: capital, licensing, accounting guidelines, promotion, goods, pricing, human resources, advertising and other activities that often hinder the business process. Financial Management for MSME Empowerment is one of the business processes run by the Ministry of Finance in improving the performance of MSMEs through the provision of business capital. But in practice, there are obstacles in the process of testing the feasibility classification of the assistance activities carried out, because it still uses inefficient manual methods. In order for the testing to be more effective and practical, integrated website-based software is needed. To achieve this, this research was conducted with the aim of producing information related to the eligibility status of MSMEs in all regions of Indonesia that are eligible or not eligible to receive financial assistance from the government. In order for the MSME eligibility status information obtained to be useful for regional offices throughout Indonesia, the algorithms used in this research are the Naive Bayes algorithm and the C4.5 algorithm. The results showed that both algorithms can be applied well in determining the eligibility of MSME assistance. The accuracy of the C4.5 algorithm is 90% while the accuracy of the Naive Bayes method is 70%. The C4.5 algorithm performs slightly better than Naive Bayes in this classification setting. The accuracy findings in this study can be compared to previous research already conducted using the same algorithm or with similar data sets. This helps determine if this research methodology results in a higher accuracy rate than previous studies.

Keywords: UMKM; Website; Classification; Naive Bayes; C4.5

1. PENDAHULUAN

Perkembangan perekonomian Indonesia melambat akibat meningkatnya inflasi. Pengusaha sering menghadapi beberapa masalah dalam memulai suatu kegiatan kewirausahaan, termasuk: modal, perizinan, akuntansi panduan, promosi, barang, harga, sumber daya manusia, periklanan dan kegiatan lain yang sering menghambat proses bisnis[1]. Tindakan pemilik usaha kecil yang kemudian disebut UMKM merupakan kegiatan usaha yang kuat akan terjadinya suatu kegagalan usaha[2]. UMKM sangat penting bagi perekonomian negara. UMKM merupakan komponen yang cukup besar dalam angkatan kerja dan dapat menstimulasi pertumbuhan ekonomi regional[3]. Namun di sisi lain UMKM memiliki kendala utama dalam mendapatkan pembiayaan berupa pembayaran kredit untuk memiliki modal usaha[4]. Pemberdayaan UMKM merupakan salah satu cara untuk meningkatkan perekonomian masyarakat Indonesia[5]. Untuk meningkatkan kinerja UMKM, Kementerian Keuangan telah melakukan berbagai kegiatan pemberdayaan seperti pelatihan, pendampingan dan pemberian modal usaha[6]. Namun dalam praktiknya, terdapat kendala dalam melakukan pemantauan dan klasifikasi kelayakan terhadap kegiatan pemberian bantuan yang dilakukan. Monitoring dan klasifikasi kelayakan yang dilakukan masih menggunakan cara manual yang seringkali tidak efisien dan tidak akurat.

Untuk mengatasi kendala tersebut, Kementerian Keuangan berencana untuk membuat perangkat lunak. Agar perangkat lunak yang dirancang dapat diakses oleh seluruh kantor wilayah Indonesia maka sistem yang dibangun berbasis website yang mampu digunakan untuk melakukan klasifikasi kelayakan kegiatan pemberian bantuan UMKM. Sistem yang dirancang menggunakan metode *Agile model Extreme Programming*, yang memungkinkan proses pengembangan dilakukan dengan cepat dan fleksibel[7]. Sistem Informasi ini diharapkan dapat membantu lebih cepat dan efektif bagi

Kantor Wilayah di seluruh Indonesia dalam memantau dan mengevaluasi kegiatan bantuan kepada UMKM, sehingga dapat meningkatkan kinerja UMKM dan berdampak positif pada perekonomian masyarakat di seluruh

Indonesia.

Metode *Agile* model *extreme programming* dalam desain dan pelaksanaan klasifikasi kelayakan bantuan UMKM adalah perbedaan utama antara penelitian ini dan penelitian sebelumnya. Penelitian sebelumnya memiliki kecenderungan hanya mengandalkan algoritma daripada teknik desain sistem. Dengan menggabungkan konsep *Agile* dengan Algoritma yang digunakan perancangan dalam pengembangan sistem, yang dapat memberikan solusi yang lebih fleksibel, efisien, dan terukur, penelitian ini berusaha untuk menutup kesenjangan ini.

Klasifikasi adalah tindakan mengklasifikasikan melibatkan pengelompokan item ke dalam kelompok yang berbeda. Kegiatan mengklasifikasikan kelayakan telah banyak dalam penelitian-penelitian sebelumnya. Pertama, "Klasifikasi dan Monitoring Status Gizi Balita Melalui Penerapan Metode Naive Bayes Classification Berbasis GIS"[8], Selanjutnya "aplikasi monitoring dan penentuan peringkat kelas menggunakan naive bayes classifier"[9]. Selanjutnya, "Penerapan algoritma C4.5 pada klasifikasi potensi siswa drop out"[10], Dan selanjutnya "Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama)"[11]. Berdasarkan penelitian yang sudah ada tersebut terdapat beberapa perbedaan antara algoritma Naive Bayes dan algoritma C4.5 dalam mengklasifikasi.

Pentingnya keterbaruan dari penelitian yang sudah ada sebelumnya adalah untuk memastikan bahwa peneliti memberikan kontribusi yang berarti, dan penelitian ini memiliki dampak positif yang lebih luas bagi masyarakat atau industri terkait dalam hal klasifikasi kelayakan bantuan UMKM, selain itu terdapat kelebihan dari segi manfaat algoritma Naive Bayes dan algoritma C4.5 yang mampu membuat efisien dalam hal pelatihan dan pengujian, serta membantu mengidentifikasi fitur atau variabel yang paling berpengaruh dalam klasifikasi kelayakan bantuan UMKM. Terdapat pula kelebihan implementasi pada algoritma Naive Bayes yang mampu mengelompokkan kategori-kategori yang ada menggunakan probabilitas dan teorema Bayes, sedangkan algoritma C4.5 mampu mengatasi data yang hilang. Daripada itu keterbaruan dapat membantu dalam pengembangan model yang lebih efektif dan aplikatif untuk membantu dalam pengambilan keputusan tentang pemberian bantuan UMKM yang lebih akurat dan tepat sasaran.

Menurut Thomas Bayes 1763 (dalam (Yang, 2019)) pengklasifikasi Naive Bayes adalah jenis mekanisme klasifikasi probabilistik berdasarkan teorema Bayes, teori anumerta Thomas Bayes. Dari sudut pandang klasifikasi, tujuan utamanya adalah menemukan deskripsi terbaik antara data baru dan klasifikasi yang ditempatkan di area masalah tertentu. Untuk membuat pemetaan ini dapat dihitung secara probabilistik, beberapa manipulasi matematis dilakukan untuk mengubahnya[12]. Sedangkan menurut Hasil dari C4.5 adalah Pohon keputusan, teknik yang sangat efektif dan terkenal untuk kategorisasi dan prediksi adalah pohon keputusan. Fakta-fakta berskala besar dikonversi menggunakan pendekatan pohon keputusan yang mewujudkan aturan. Aturan-aturan tersebut mudah dipahami ketika diucapkan dalam bahasa normal. Dan untuk mencari entri dalam kategori tertentu, mereka juga dapat dijelaskan menggunakan bahasa database seperti *Structured Query Language*. [13]

Diharapkan dengan adanya sistem informasi pengklasifikasian kelayakan bantuan UMKM pada Kanwil Ditjen Perbendaharaan ini akan menghasilkan sebuah sistem yang dapat menyajikan informasi mengenai UMKM yang layak, maupun tidak layak menerima bantuan pemerintah, sehingga pemerintah mampu melakukan pengontrolan terhadap kelayakan UMKM dan dapat lebih waspada agar tidak melakukan kesalahan pada saat memberikan bantuan kepada UMKM.

2. METODOLOGI PENELITIAN

Penelitian dilakukan untuk memperoleh fakta-fakta yang dibutuhkan, maka dilakukan penelitian pada peneliti melakukan penelitian dalam jangka waktu kurang lebih 4 bulan, yakni Maret 2023 - Juni 2023 di Direktorat Jenderal Perbendaharaan yang beralamatkan di Gedung Direktorat Jenderal Perbendaharaan Lt. 3, Jl. Lapangan Banteng Timur, Jakarta Pusat. Penggunaan beberapa pada klasifikasi kelayakan ini seperti, Nama, NIK (Nomor Induk Keluarga), Alamat (Provinsi, Kota/Kabupaten, Kecamatan, Kode Pos), *E-mail*, Nomor Telepon / Nomor Handphone, Nama Merk, Bidang Usaha, NIB (Nomor Izin Berusaha), PKH (Program Keluarga Harapan), Jumlah Tanggungan, Kondisi Rumah, Jumlah Penghasilan, Jenis Kelamin, Status Rumah. Sistem ini digunakan oleh satu pengguna yaitu, administrator atau pengguna.

2.1 Sumber dan Cara Pengumpulan Data

Peneliti memanfaatkan sumber data primer dan sekunder untuk mendapatkan informasi

- Data ini mengumpulkan langsung dari subjek penelitian serta penelitian lapangan langsung untuk proses pemecahan masalah yang dipelajari oleh peneliti.
- Sumber data sekunder dikumpulkan oleh peneliti untuk mendukung sumber pertama, dapat dikatakan bahwa informasi yang diperoleh dari jurnal – jurnal, artikel, dokumen yang dianggap relevan sehingga bisa untuk memperkuat proses penelitian.

Dalam rangka mengumpulkan data untuk penelitian ini, wawancara dan dokumentasi digunakan didapat dari tempat penelitian, adapun penjelasannya adalah :

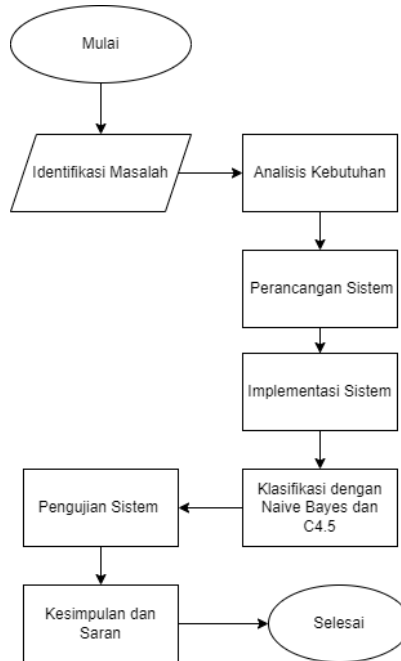
- Wawancara

Wawancara, sebuah prosedur tanya-jawab langsung antara peneliti dan narasumber, merupakan suatu metode untuk pengumpulan data dalam penelitian ini. Wawancara ini dilakukan dengan pihak dari Direktor Jenderal Perbendaharaan (Direktorat Sistem Manajemen Investasi) sebagai narasumber, yang bertujuan untuk memudahkan penulis mendapatkan informasi yang dibutuhkan.

b. Dokumentasi

Dokumentasi adalah metode pengumpulan informasi melalui dokumen atau catatan yang ada. Metode dokumen digunakan untuk mencari informasi yang ada, yaitu dokumen tentang orang atau kelompok dan juga tentang peristiwa di lingkungan sosial. Dokumen yang penulis gunakan dalam proses pengumpulan data adalah formulir penelitian untuk survey.

2.2 Tahapan Penelitian



Gambar 1. Tahapan Penelitian

a. Identifikasi Masalah

Mengenali masalah yang dipelajari selama langkah identifikasi[14]. Ini adalah tahap penting dalam klasifikasi kelayakan bantuan untuk UMKM, pendistribusian bantuan untuk UMKM di Kantor Wilayah Direktorat Jenderal Perbendaharaan yang belum akurat sehingga membuat proses penyaluran bantuan terhambat.

b. Analisis Kebutuhan

Dalam tahap ini proses pengumpulan informasi mengenai data-data yang dibutuhkan dalam pembuatan sistem klasifikasi kelayakan bantuan UMKM[15]. Pada pengumpulan data ini penulis melakukan wawancara dan juga studi dokumentasi untuk mencari informasi beserta data mengenai UMKM.

Tabel 1. Jenis Data Pendataan UMKM

Nama Data	Keterangan
Nama Pemilik	Berisikan nama dari pemilik UMKM
NIK	Berisikan Nomor Induk Kependudukan
Alamat	Berisikan data alamat rumah pemilik usaha (Provinsi, Kota/Kabupaten, Kecamatan, Kode pos)
E-mail	Akun E-mail yang dimiliki oleh pemilik UMKM
Nomor Telepon/Nomor Handphone	Nomor Telepon/Nomor Handphone yang dimiliki oleh pemilik UMKM
Nama Merek Bidang Usaha	Nama merek dari usaha yang dijalankan
NIB	Bidang usaha apa yang dijalankan oleh pemilik usaha (Fashion, Makanan, Minuman, Barang)
	Nomor Izin Berusaha yang didapatkan oleh pemilik usaha.

Tabel 2. Daftar Data Klasifikasi

Nama Data	Kategori	Keterangan
PKH	PKH	Status Program Keluarga Harapan (PKH /Non PKH) dari pemilik usaha tersebut mendapatkan bantuan PKH atau Tidak
	Non PKH	
Jumlah Tanggungan	0-5	Jumlah tanggungan yang dinafkahi
	5-10	

Jenis Kelamin	Laki-laki Perempuan	Status jenis kelamin dari pemilik UMKM
Kondisi Rumah	Batu Permanen; Batu Anyam; Papan	Kondisi rumah yang ditempati
Jumlah Penghasilan	>2.500.000 (besar) >1.500.000 – 2.500.000 (sedang) <1/500/000 (kecil)	Jumlah penghasilan yang didapatkan dari pemilik UMKM selama 1 bulan
Status Pemilik Rumah	Milik Sendiri; Sewa	Status kepemilikan rumah (Milik Sendiri atau Sewa)

c. Perancangan Sistem

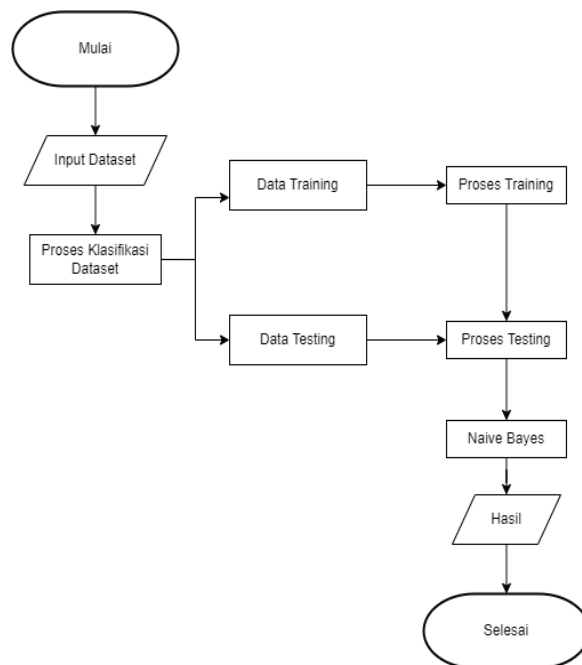
Perancangan sistem dilakukan pada tahap ini dengan menggunakan komponen-komponen *UML (Unified Modeling Language)* seperti *flowchart*, *use case diagram*, dan *activity diagram*[16].

d. Implementasi Sistem

Pada tahap ini melakukan penulisan program untuk membuat sistem yang berbasis website, penulisan program untuk membuat CRUD (Menambahkan, Melihat, Memperbarui, dan Menghapus) data, penggabungan algoritma Naive Bayes dan C4.5 ke dalam bahasa pemrograman PHP, serta spesifikasi perangkat keras dan perangkat lunak sistem.[17]

e. Klasifikasi

Tahap Klasifikasi ini dimaksudkan untuk melakukan perhitungan secara manual ataupun memanfaatkan algoritma C4.5 dan algoritma Naive Bayes di Rapidminer.



Gambar 2. Flowchart Algoritma Naïve Bayes

Gambar 2. merupakan alur tahapan dalam implementasi algoritma Naïve Bayes untuk mengklasifikasi calon penerima bantuan keuangan.

1. Pengambilan dari data pengujian yaitu data *training*
2. Setelah menentukan data *training* selanjutnya, memasukkan data *testing*.
3. Setelah mengubah kategori tiap atribut pada data testing yang sudah dimasukkan, dan melakukan perhitungan probabilitas.
4. Dapat hasil prediksi

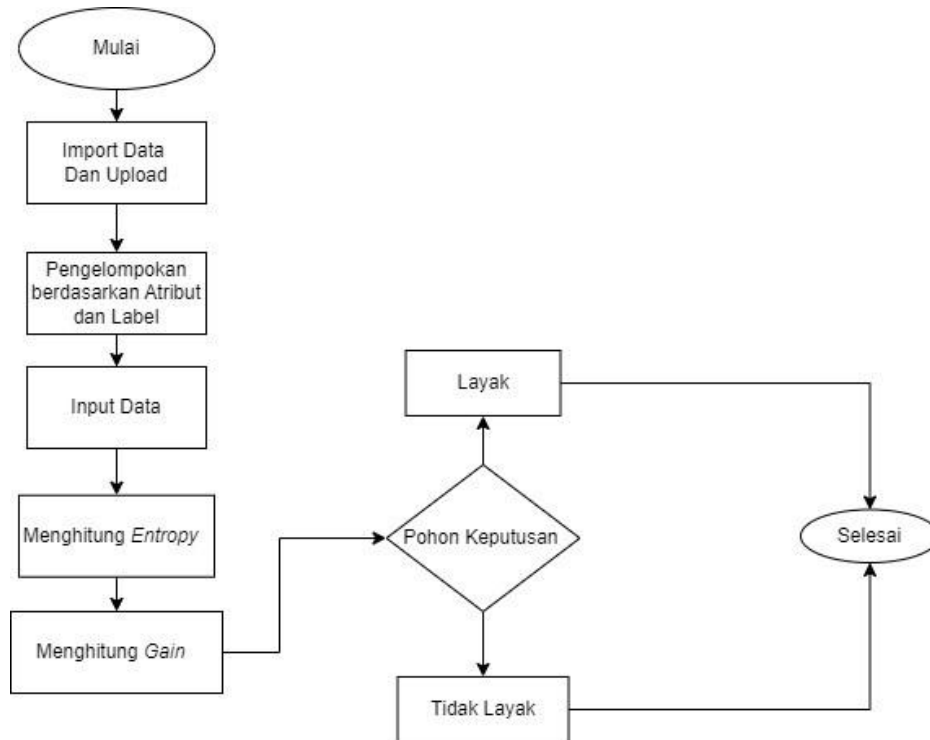
Perhitungan algoritma Naïve Bayes, dengan menggunakan rumus berikut:

- X = Data dengan *class* yang belum diketahui
- H = Hipotesis data X merupakan suatu *class* spesifik
- $P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (*Posteriori Probability*)
- $P(H)$ = Probabilitas hipotesis H (*Prior Probability*)
- $P(X|H)$ = Probabilitas X berdasarkan hipotesis
- $HP(X)$ = Probabilitas dari X [18]

Naïve Bayes merupakan penyederhanaan metode bayes. Teorema bayes disederhanakan menjadi:

$$P(H|X) = P(X|H)P(X) \quad (2)$$

Metode C4.5 memiliki beberapa fase untuk membangun pohon keputusan:



Gambar 3. Flowchart Algoritma C4.5

1. Program uji kelayakan diawali dengan *login* dari administrator sebelum berlanjut ke proses *import* data berupa data dalam format excel.
2. Sistem informasi juga akan menganalisis data yang di *import* sehingga dapat dihasilkan dalam bentuk tabel dengan format yang telah ditentukan.
3. Dari hasil *import* data, selanjutnya sistem informasi akan memproses pengelompokan menjadi warna kuning (Atribut) dan hijau (Label).
4. Selanjutnya yaitu melakukan *input* data untuk di prediksi kelayakannya.
5. Menghitung *Entropy* merupakan menghitung setiap atribut yang ada.

Gunakan rumus berikut untuk menghitung nilai entropy:

$$Entropy(S, A) = entropy(S) \sum_{i=1}^n - p_i * \log_2 p_i \quad (3)$$

Keterangan:

S = Himpunan

Kasus = Jumlah partisi S

p_i = Proporsi S_i terhadap S [19]

6. Menghitung *Gain* merupakan menghitung total hasil setiap *Entropy* pada atribut.

$$Gain(S, A) = entropy(S) \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (4)$$

Keterangan

S = Himpunan Kasus

A = Fitur

N = Jumlah partisi atribut A

$|S_i|$ = Proporsi S_i terhadap S

$|S|$ = Jumlah kasus dalam S [19]

7. Selanjutnya baru didapatkan pohon keputusan dengan cara total *Gain* yang tertinggi dari masing-masing atribut yang ada.

f. Pengujian Sistem

Fase pengujian sistem perangkat lunak sangat penting. Sistem diuji pada tahap ini untuk memastikan sistem dapat digunakan dengan baik, tanpa kesalahan, dan sesuai permintaan dari pengguna. Pengujian *Black Box* digunakan dalam Pengujian Sistem ini [20]. Apakah sistem atau program menghasilkan hasil yang diinginkan dalam menanggapi *input* yang bervariasi tergantung pada kita.

3. HASIL DAN PEMBAHASAN

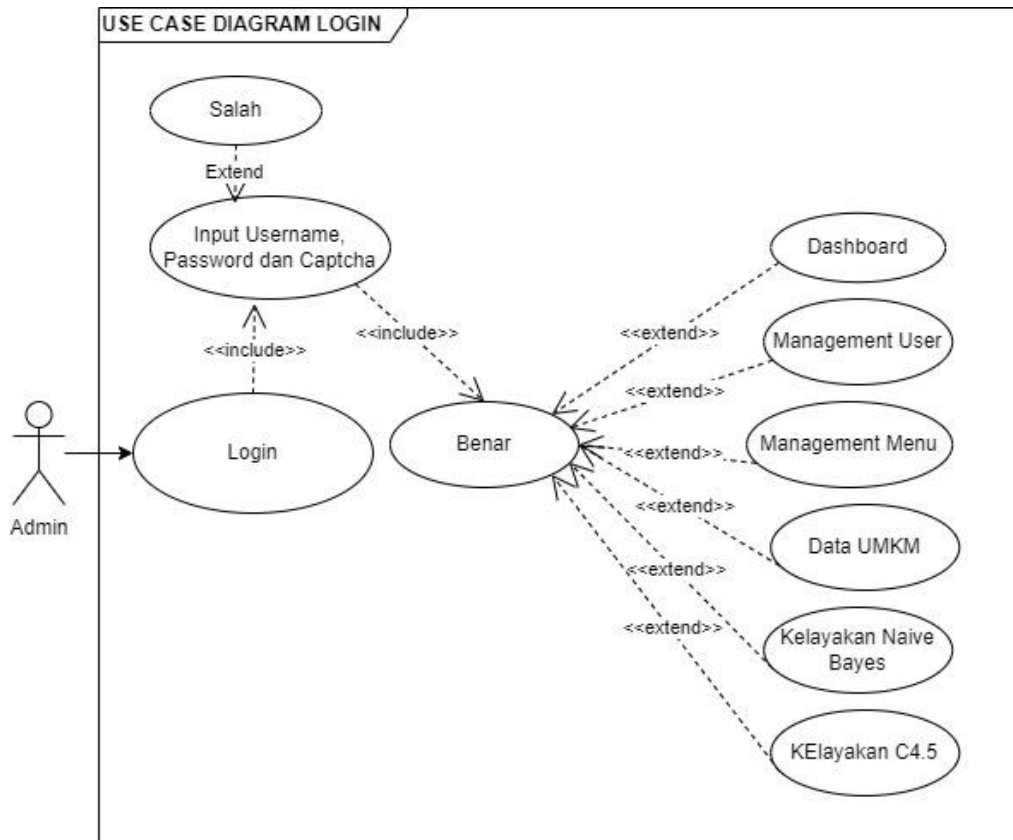
3.1 Pengumpulan Data

Pengumpulan data meliputi wawancara dan dokumentasi yang dilakukan di Direktorat Jenderal Perbendaharaan, Kementerian Keuangan. Data pelatihan yang digunakan untuk membangun sistem klasifikasi kelayakan bantuan UMKM dengan menggunakan algoritma Nave Bayes dan C4.5 ditunjukkan pada tabel di bawah ini. Data set ini berisi sejumlah atribut yang berkaitan dengan karakteristik UMKM.

Tabel 3. Data Training

Nama	PKH	Jumlah Tanggungan	Jenis Kelamin	Kondisi Rumah	Jumlah Penghasilan	us Pemilik Rumah	Status Kelayakan
Aldy	PKH	5	Laki-laki	Bambu anyam	400000	Milik sendiri	Layak
Budi	Non-PKH	3	Laki-laki	Papan	3000000	Sewa	Tidak layak
Citra	PKH Non-	3	Laki-laki	Bambu anyam	3000000	Milik Sendiri	Tidak Layak
Dewi	PKH	1	Laki-laki	Bambu anyam	300000	Sewa	Tidak Layak
Eka	PKH PKH	3	Perempuan	Bambu anyam	3000000	Milik Sendiri	Tidak Layak
Fajar	PKH	3	Laki-laki	Bambu anyam	3000000	Milik Sendiri	Tidak Layak
Gita	Non-PKH	3	Perempuan	Batu Permanen	2000000	Milik Sendiri	Tidak Layak
Hadi	PKH	3	Laki-laki	Batu Permanen	300000	Sewa	Layak
Indra	Non-PKH	3	Laki-laki	Bambu anyam	3000000	Milik Sendiri	Tidak Layak
Ibu		3	Perempuan	Batu Permanen	200000	Sewa	Layak

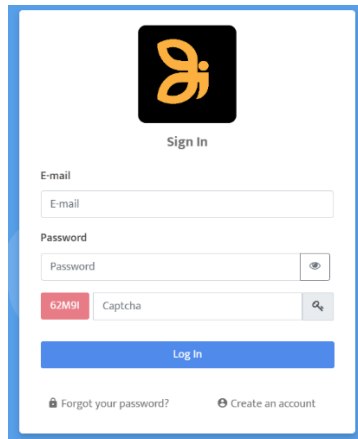
3.2 Perancangan Sistem



Gambar 4. Use Case Diagram Login

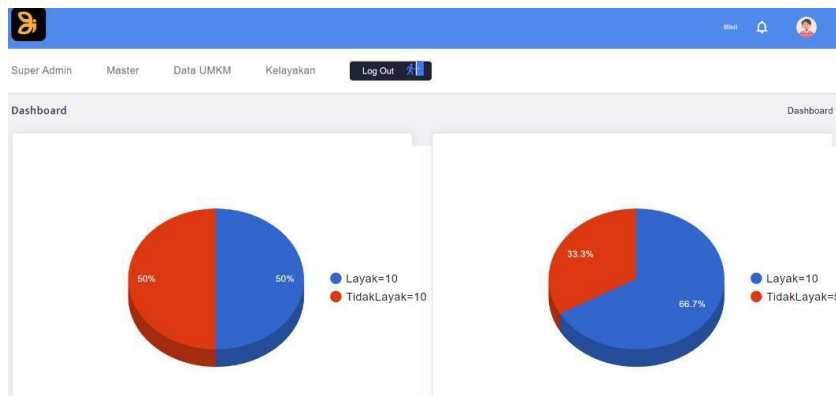
Use Case Diagram Login tersebut menunjukkan bahwa Admin harus memasukkan E-mail, Password, dan Captcha terlebih dahulu agar bisa Login. Ketika sudah benar Admin akan di lanjutkan ke Halaman Dashboard yang di dalamnya terdapat Pie Chart.

3.3 Implementasi Sistem



Gambar 5. Tampilan Login

Menunjukkan tampilan login yang digunakan oleh Administrator untuk masuk kedalam menu *dashboard*. Admin akan mengakses halaman *dashboard* setelah memasukkan alamat email dan kata sandi secara akurat. Jika salah memasukkannya, maka admin akan mendapatkan pesan atau tanda bahwa gagal masuk.



Gambar 6. Tampilan Dashboard

Tampilan *Dashboard* adalah tampilan yang muncul ketika administrator sudah berhasil *login*. Pada Tampilan ini Admin bisa memonitoring Status Kelayakan, Status Rumah, Status Kepemilikan Rumah, dan PKH (ProgramKeluarga Harapan) semuanya di tampilkan dalam bentuk *Pie Chart*. Menu-menu yang dipilih untuk situs web lengkap yang didukung oleh aplikasi disertakan dalam dashboard. Menu-menu utama tersebut antara lain SuperAdmin, menu master, Data UMKM, Kelayakan, dan Kelayakan C4.5.

3.4 Klasifikasi Naïve Bayes dan C4.5

Tabel 4. Data Testing

Nama	PKH	Jumlah Tanggungan	Jenis Kelamin	Kondisi Rumah	Jumlah Penghasilan	Status PemilikRumah	Status Kelayakan
Jaka	Non-PKH	3	Laki-laki	Bambu Permanen	30000000	Milik sendiri	?

Dengan rumus Naïve Bayes dapat dihitung probabilitasnya:

- a. Menghitung jumlah variable kelas/label

$$P(Y=Layak) = 3/10 = 0,3$$

$$P(Y=Tidak Layak) = 7/10 = 0,7$$

- b. Menjumlahkan kasus dengan kelas yang sama:

$$P(PKH = Non-PKH | Y=Layak) = 2/3 = 0,6$$

$$P(PKH = Non-PKH | Y=Tidak Layak) = 2/7 = 0,28$$

$$P(\text{Jumlah Tanggungan} = 0-5 | Y=Layak) = 3/3 = 1 \quad P(\text{Jumlah Tanggungan} = 0-5 | Y=Tidak Layak) = 7/7 = 1$$

$$P(\text{Jenis Kelamin} = L | Y=Layak) = 2/3 = 0,66 \quad P(\text{Jenis Kelamin} = L | Y=Tidak Layak) = 5/7 = 0,7$$

$$P(\text{Kondisi Rumah} = Batu Permanen | Y=Layak) = 2/3 = 0,66 \quad P(\text{Kondisi Rumah} = Batu Permanen | Y=Tidak Layak) = 1/7 = 0,14$$

$$P(\text{Jumlah Penghasilan} = >2.500.000 | Y=Layak) = 0/3 = 0 \quad P(\text{Jumlah Penghasilan} = >2.500.000 | Y=Tidak Layak) = 5/7 = 0,7$$

$P(\text{Status Pemilik Rumah} = \text{Milik Sendiri} \mid Y=\text{Layak}) = 1/3 = 0,3$
 $P(\text{Status Pemilik Rumah} = \text{Milik Sendiri} \mid Y=\text{Tidak Layak}) = 5/7 = 0,7$
 Kalian semua hasil label layak dan tidak layak
 $P(\text{PKH} = \text{Non-PKH} \mid Y=\text{Layak}) \times P(\text{Jumlah Tanggungan} = 0-5 \mid Y=\text{Layak}) \times P(\text{Jenis Kelamin} = L \mid Y=\text{Layak}) \times P(\text{Kondisi Rumah} = \text{Batu Permanen} \mid Y=\text{Layak}) \times P(\text{Jumlah Penghasilan} = >2.500.000 \mid Y=\text{Layak}) \times P(\text{Status Pemilik Rumah} = \text{Milik Sendiri} \mid Y=\text{Layak}) \times P(Y=\text{Layak}) = 0,6 \times 1 \times 0,66 \times 0,66 \times 0 \times 0,3 \times 0,3 = 0$
 $P(\text{PKH} = \text{Non-PKH} \mid Y=\text{Tidak Layak}) \times P(\text{Jumlah Tanggungan} = 0-5 \mid Y=\text{Tidak Layak}) \times P(\text{Jenis Kelamin} = L \mid Y=\text{Tidak Layak}) \times P(\text{Kondisi Rumah} = \text{Batu Permanen} \mid Y=\text{Tidak Layak}) \times P(\text{Jumlah Penghasilan} = >2.500.000 \mid Y=\text{Tidak Layak}) \times P(\text{Status Pemilik Rumah} = \text{Milik Sendiri} \mid Y=\text{Tidak Layak}) \times P(Y=\text{Tidak Layak}) = 0,28 \times 1 \times 0,7 \times 0,14 \times 0,7 \times 0,7 \times 0,7 = 0,00941192$

- c. Bandingkan antara hasil yang di dapatkan
 Dari hasil layak dan tidak layak diatas, nilai probabilitas tertinggi terletak pada label tidak layak, jadi dapat disimpulkan Pemilik UMKM tersebut tidak layak untuk mendapatkan bantuan dari Pemerintah.

Perhitungan Manual dengan Algoritma C4.5:

- a. Hitung *Entropy* Awal:
 Jumlah total *entri* yang layak :3
 Jumlah total *entri* yang tidak layak :7
 Total *entri* dalam *dataset*:10
 $Entropy \text{ Awal} = -((3/10) * \log_2(3/10)) - ((7/10) * \log_2(7/10)) = 0.8812$
- b. *Gain* untuk Atribut PKH:
Entropy setelah *splitting* pada nilai PKH:
 Untuk nilai PKH: 6 entri (1 layak, 5 tidak layak)
 Untuk nilai Non-PKH: 4 entri (2 layak, 2 tidak layak)
 $Entropy(\text{PKH}) = -((1/6) * \log_2(1/6)) - ((5/6) * \log_2(5/6)) = 0.6500$
 $Entropy(\text{Non-PKH}) = -((2/4) * \log_2(2/4)) - ((2/4) * \log_2(2/4)) = 0.5$
 $Gain \text{ Total} = 0.8812 - ((6/10) * 0.6500) - ((4/10) * 0.5) = 0.2912$
- c. *Gain* untuk Atribut Jumlah tanggungan:
Entropy setelah *splitting* pada nilai Jumlah tanggungan:
 Untuk nilai Jumlah Tanggungan ≤ 2 : 1 *entri* (0 layak, 1 tidak layak)
 Untuk nilai Jumlah Tanggungan > 2 : 9 *entri* (3 layak, 6 tidak layak)
 $Entropy(\leq 2) = -((0/1) * \log_2(0/1)) - ((1/1) * \log_2(1/1)) = 0$
 $Entropy(> 2) = -((3/9) * \log_2(3/9)) - ((6/9) * \log_2(6/9)) = 0.9182$
 $Gain \text{ Total} = 0.8812 - ((1/10) * 0) - ((9/10) * 0.9182) = 0.05482$
- d. *Gain* untuk Atribut Kondisi rumah:
Entropy setelah *splitting* pada nilai Kondisi rumah:
 Untuk nilai Papan: 1 *entri* (0 layak, 1 tidak layak)
 Untuk nilai Batu permanen: 3 *entri* (2 layak, 1 tidak layak)
 Untuk nilai Bambu anyam: 6 *entri* (1 layak, 5 tidak layak)
 $Entropy(\text{Papan}) = -((0/1) * \log_2(0/1)) - ((1/1) * \log_2(1/1)) = 0$
 $Entropy(\text{Batu permanen}) = -((2/3) * \log_2(2/3)) - ((1/3) * \log_2(1/3)) = 0.9182$
 $Entropy(\text{Bambu anyam}) = -((1/6) * \log_2(1/6)) - ((5/6) * \log_2(5/6)) = 0.6500$
 $Gain \text{ Total}(\text{Kondisi rumah}) = 0.8812 - ((1/10) * 0) - ((3/10) * 0.9182) - ((6/10) * 0.6500) = 0.21574$
- e. *Gain* untuk Atribut Jumlah penghasilan:
Entropy setelah *splitting* pada nilai Jumlah penghasilan:
 Untuk nilai ≤ 1500000 : 4 *entri* (3 layak, 1 tidak layak)
 Untuk nilai > 1500000 : 6 *entri* (0 layak, 6 tidak layak)
 $Entropy(\leq 1500000) = -((3/3) * \log_2(3/3)) - ((1/3) * \log_2(1/3)) = 0.5283$
 $Entropy(> 1500000) = -((0/6) * \log_2(0/6)) - ((6/6) * \log_2(6/6)) = 0$
 $Gain \text{ Total}(\text{Jumlah penghasilan}) = 0.8812 - ((4/10) * 0.5283) - ((6/10) * 0) = 0.66988$
- f. *Gain* untuk Atribut Status kepemilikan:
Entropy setelah *splitting* pada nilai Status kepemilikan:
 Untuk nilai Sewa: 4 *entri* (2 layak, 2 tidak layak)
 Untuk nilai Milik sendiri : 6 *entri* (1 layak, 5 tidak layak)
 $Entropy(\text{Sewa}) = -((2/4) * \log_2(2/4)) - ((2/4) * \log_2(2/4)) = 1$
 $Entropy(\text{Milik sendiri}) = -((1/6) * \log_2(1/6)) - ((5/6) * \log_2(5/6)) = 0.6500$
 $Gain \text{ Total}(\text{Status kepemilikan}) = 0.8812 - ((4/10) * 1) - ((6/10) * 0.6500) = 0.0912$

Tabel 5. Hasil Klasifikasi C4.5

Kasus	Kategori	Jumlah Layak	Tidak Layak	Entropy	Gain	
PHK	PKH Non-PKH	10	3	7	0.8812	0.2912
		6	1	5	0.6500	
		4	2	2	0.5	

Jumlah Tanggungan	<=2	1	0	1	0	05482
	>2	9	3	6	0.2912	
KondisiRumah	Batu Permanen	3	2	1	0.9182	0.21574
	Batu Anyam	6	1	5	0.6500	
	Papan	1	0	1	0	
Jumlah Penghasilan	<=1500000	4	3	1	0.5283	0.66988
	>1500000	6	0	6	0	
Status Kepemilikan Rumah	ik Sendiri	6	1	5	0.6500	0.09212
	Sewa	4	2	2	1	

Hasil perhitungan diatas, disimpulkan bahwa jumlah penghasilan merupakan faktor paling kuat yang mempengaruhi apakah penerima bantuan UMKM layak atau tidak., karena memiliki nilai *Gain* paling tinggi.

Perhitungan menggunakan RapidMiner di dapatkan hasil *accuracy*, *precision*, dan *recall*, sebagai berikut:

Tabel 6. Hasil RapidMiner Algoritma Naïve Bayes

Evaluasi	Persentase
<i>Accuracy</i>	70.00%
<i>Precision</i>	75.00%
<i>Recall</i>	85.71%

Tabel 7. Hasil RapidMiner Algoritma C4.5

Evaluasi	Persentase
<i>Accuracy</i>	90.00%
<i>Precision</i>	100.00%
<i>Recall</i>	85.71%

Hasil pengujian menggunakan RapidMiner menunjukkan algoritma C4.5 mampu memberikan hasil yang lebih baik bila dibandingkan dengan algoritma Naïve Bayes. Namun untuk *Recall* mendapatkan nilai persentase yang sama di antar kedua algoritma tersebut.

3.6 Pengujian Sistem

Hasil pengujian sistem ini menggunakan metode *BlackBox* dianalisis untuk menemukan hasil pengujian terperinci untuk setiap sistem halaman yang diuji. Data yang digunakan untuk melakukan pengujian ini menggunakan data *dummy* atau data tidak benar. Data yang digunakan seperti Tabel 1 dan Tabel 2, dibawah ini merupakan hasil pengujian metode *Blackbox*:

Tabel 8. Hasil Pengujian *Black Box*

Fitur	Scenario Fitur	Output yang Dihasilkan	Kesimpulan
<i>Login</i>	Memasukkan <i>e-mail</i> , <i>password</i> , dan <i>Code Captcha</i>	Berhasil melakukan <i>Login</i>	[x] Terima [] Tolak
PendataanUMKM	Melakukan input dataUMKM	Menampilkan data UMKM	[x] Terima [] Tolak
<i>Searching</i>	Melakukan <i>Searching</i> data disetiap menu	Menampilkan hasil data sesuai <i>keyword</i> yang dimasukkan di <i>field search</i> data	[x] Terima [] Tolak
Kelayakan (Calon)	Melakukan input data calon yang ingin di lakukan pengujian	Menampilkan data calon	[x] Terima [] Tolak
Kelayakan Pengujian	Melakukan <i>input</i> dataUMKM yang ingin di lakukan pengujian	Menampilkan hasil pengujian layak atau tidak layak	[x] Terima [] Tolak
Export daftar Data Pengujian	Meng <i>export</i> daftar data pengujian	Data berhasil terunduh dalam bentuk Excel	[x] Terima [] Tolak
Kelayakan C4.5	Melakukan <i>upload</i> data yang sudah di unduh.	Data berhasil di <i>upload</i>	[x] Terima [] Tolak
Kelayakan C4.5	Melakukan input data UMKM yang di lakukan Pengujian	Menampilkan hasil pengujian kelayakan	[x] Terima [] Tolak

Dari hasil pengujian yang dilakukan oleh pegawai kantor wilayah Direktorat Jenderal Perbendaharaan, dapat ditarik kesimpulan bahwa fitur-fitur yang ada diterima oleh pengguna, berjalan dengan lancar sistem informasi yang digunakan, mampu melakukan klasifikasi kelayakan sesuai dengan yang diinginkan oleh pengguna.

4. KESIMPULAN

Kesimpulan yang dapat diambil dari temuan-temuan penilaian kelayakan bantuan dengan memakai algoritma Naive Bayes dan C4.5 adalah sistem informasi ini dapat memudahkan pekerjaan dan menjadi lebih efisien dalam hal menginputkan data dan pengujian kelayakan bantuan UMKM, selain itu Data UMKM yang relevan digunakan sebagai input dalam algoritma Naive Bayes dan C4.5 yang kemudian dilakukan uji klasifikasi menentukan kelayakan bantuan UMKM. Berdasarkan hasil pengujian algoritma Naive Bayes dan C4.5 mampu memberikan hasil klasifikasi yang baik dalam mengklasifikasikan kelayakan bantuan UMKM. Dan Hasil pengujian menunjukkan algoritma C4.5 mampu bersaing dengan hasil yang lebih baik bila dibandingkan dengan algoritma Naive Bayes. *Accuracy* algoritma C4.5 sebesar 90.00% dan 70.00% Algoritma Naive Bayes. Tetapi kedua Algoritma mampu melakukan klasifikasi kelayakan bantuan UMKM dengan baik. Namun, ada beberapa kendala yang perlu diatasi dalam implementasi sistem informasi klasifikasi kelayakan bantuan UMKM kedepannya seperti masalah keamanan data, keterbatasan infrastruktur, dan pengumpulan data yang lebih banyak. Oleh karena itu, perlu pengembangan serta perbaikan perangkat secara terus-menerus untuk memastikan keberhasilan penggunaan sistem ini dalam meningkatkan klasifikasi kelayakan bantuan pada UMKM.

REFERENCES

- [1] T. Ardiansyah, "Model Financial Dan Teknologi (Fintech) Membantu Permasalahan Modal Wirausaha UMKM Di Indonesia," *Jurnal Majalah Ilmiah Bijak*, vol. 16, no. 2, pp. 158-166, Sep. 2019, doi: 10.31334/bijak.v16i2.518.
- [2] L. M. Hamza and D. Agustien, "Pengaruh Perkembangan Usaha Mikro, Kecil, dan Menengah Terhadap Pendapatan Nasional Pada Sektor UMKM di Indonesia," *Jurnal Ekonomi Pembangunan*, vol. 8, no. 2, pp. 127-135, Jul. 2019, doi: 10.23960/jep.v8i2.45.
- [3] L. Marlinah, "Peluang dan Tantangan UMKM Dalam Upaya Memperkuat Perekonomian Nasional Tahun 2020 Ditengah Pandemi Covid 19," *Jurnal Ekonomi*, vol. 22, no. 2, pp. 118-124, Jun. 2020, doi: 10.37721/je.v22i2.644.
- [4] T. Fitri, P. Dewi, Y. Remofa, and) Hardi, "STRATEGI UMKM AKSES PENDANAAN PERBANKAN DI DESA SEMELINANG DARAT PERANAP," *Jurnal Pengabdian Kepada Masyarakat*, vol. 5, no. 1, pp. 18-27, Jun. 2023, doi: 10.34006/values.v5i1.543.
- [5] F. Ekonomi et al., "ANALISIS PEMBERDAYAAN UMKM PADA MASA PANDEMI COVID-19," *SOSEBI: Jurnal Penelitian Mahasiswa Ilmu Sosial, Ekonomi, dan Bisnis Islam*, vol. 83, no. 1, pp. 2808-7089, Nov. 2021, doi: 10.21274/sosebi.v1i1.4929.
- [6] D. Putri Nasution, "ANALISIS EKSISTENSI UMKM DALAM MENGURANGI TINGKAT KEMISKINAN DI INDONESIA." *Jurnal Edunomika*, vol. 7, no. 1, pp. 1-12, Agu. 2023, doi: 10.29040/jie.v7i1.7002.
- [7] F. Fatoni and D. Irawan, "Implementasi Metode Extreme Programming dalam Pengembangan Sistem Informasi Izin Produk Makanan," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 8, no. 2, pp. 159-164, Aug. 2019, doi: 10.32736/sisfokom.v8i2.679.
- [8] R. Darma Rusdiyana Yusron and I. Machfud, "Klasifikasi dan Monitoring Status Gizi Balita Melalui Penerapan Metode Naive Bayes Classification Berbasis GIS Classification and Monitoring of Toddler Nutrition Status Through Application of GIS-Based Naive Bayes Classification Method," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 4, no. 02, pp. 161-168, Nov. 2022, doi: 10.46772/intech.v4i02.869.
- [9] B. S. Gandhi, D. A. Megawaty, and D. Alita, "APLIKASI MONITORING DAN PENENTUAN PERINGKAT KELAS MENGGUNAKAN NAIVE BAYES CLASSIFIER," *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 2, no. 1, pp. 54-63, Mar. 2021, doi: 10.33365/jatika.v2i1.722.
- [10] M. Muhamad, A. P. Windarto, and S. Suhada, "PENERAPAN ALGORITMA C4.5 PADA KLASIFIKASI POTENSI SISWA DROP OUT," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, pp. 753-760, Dec. 2019, doi: 10.30865/komik.v3i1.1688.
- [11] C. Agus Sugianto and F. Rizky Maulana, "Algoritma Naive Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama) Naive Bayes Algorithm For Classification of Non-Cash Food Aid Recipients (Case Study of Utama Village)," *Techno.com*, vol. 18, no. 4, pp. 321-331, Nov. 2019, doi: 10.33633/te.v18i4.2587.
- [12] F. J. Yang, "An implementation of naive bayes classifier," in *Proceedings - 2018 International Conference on Computational Science and Computational Intelligence, CSCI 2018, Institute of Electrical and Electronics Engineers Inc.*, Dec. 2018, pp. 301-306. doi: 10.1109/CSCI46756.2018.00065.
- [13] M. Windarti and A. Suradi, "PERBANDINGAN KINERJA 6 ALGORITME KLASIFIKASI DATA MINING UNTUK PREDIKSI MASA STUDI MAHASISWA," *Jurnal Telematika*, vol. 1, no. 1, pp. 14-30, Feb. 2019, doi: 10.35671/telematika.v1i2i1.778
- [14] I. Ahmad, A. Febrian, and A. T. Prastowo, "PENERAPAN DAN PENDAMPINGAN SISTEM TRACER STUDY SECARA ONLINE PADA MA MA'ARIF 1 PUNGGUR," *Communnity Development Journal*, vol. 3, no. 1, pp. 277-282, Feb. 2022, doi: 10.31004/cdj.v3i1.3813.
- [15] R. I. Borman, A. T. Priandika, and A. R. Edison, "Implementasi Metode Pengembangan Sistem Extreme Programming (XP) pada Aplikasi Investasi Peternakan," *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 8, no. 3, p. 272, Jul. 2020, doi: 10.26418/justin.v8i3.40273.
- [16] R. R. D. Gunawan, T. Oktavia, and R. I. Borman, "PERANCANGAN SISTEM INFORMASI BEASISWA PROGRAM INDONESIA PINTAR (PIP) BERBASIS ONLINE (TUDI KASUS : SMA N 1 KOTA BUMI)," *Jurnal Mikrotik*, vol. 8, no. 1, Jul. 2018, <https://ojs.ummetro.ac.id/index.php/mikrotik/article/view/745/547>.
- [17] I. Ahmad, R. Indra Borman, J. Fakhrurozi, and G. G. Caksana, "Software Development Dengan Extreme Programming (XP) Pada Aplikasi Deteksi Kemiripan Judul Skripsi Berbasis Android," *Jurnal inovtek polbeng - seri informatika*. vol. 5, no. 2, pp. 297-307, Nov. 2020