

Klasifikasi Keluarga Sejahtera Study Kasus : Kecamatan Kota Palembang

Ilsa Palingga Ninditama^{1,*}, Ilsa Palingga Ninditama¹, Widya Cholil²,
Muhamad Akbar³, Darius Antoni⁴

¹Teknik Informatika, Universitas Bina Darma, Palembang, Indonesia
Email: ^{1,*}ilsapalingga28@gmail.com, ²widya@binadarma.ac.id, ³muhamad.akbar@binadarma.ac.id,
⁴dariusantoni@mail.binadarma.ac.id
) Email Penulis Utama

Abstrak— Salah satu faktor dominan dalam tercapainya pembangunan nasional adalah jumlah penduduk yang besar dan berkualitas. Besarnya jumlah penduduk pada Kecamatan di Kota Palembang belum menghasilkan pengetahuan yang bermanfaat untuk kecamatan-kecamatan tersebut. Penelitian ini bertujuan melakukan klasifikasi tingkat kesejahteraan keluarga menggunakan data fokus pada rentang tahun 2015 hingga 2020 dengan 10.187 record data. Rapidminer pada penelitian ini sebagai alat untuk menganalisis data dari tahap preprocessing, data pelatihan dan mengklasifikasikan variabel menggunakan *Fuzzy K-Nearest Neighbor* dan *Naïve Bayes* untuk membuat model dan menentukan kriteria yang akan membantu BKKBN dalam menentukan keluarga sejahtera guna mempermudah penyuluh lapangan KB dalam melaksanakan konseling KB. Dari hasil analisis data didapatkan bahwa penggunaan naïve bayes 93,99% lebih akurat dibandingkan k-NN dengan akurasi 87,28%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa klasifikasi keluarga sejahtera telah mencapai tujuan visi Pemerintah.

Kata Kunci: Klasifikasi, K-NN, *Naïve Bayes*, Keluarga Sejahtera, Data Mining

Abstract— *One of the dominant factors in achieving national development is a large and quality population. A large number of residents in the sub-districts in Palembang City has not produced useful knowledge for these sub-districts. This study aims to classify the level of family welfare using focused data in the 2015 to 2020 range with 10,187 data records. Rapidminer in this study as a tool to analyze data from the preprocessing stage, training data and classifying variables using Fuzzy K-Nearest Neighbor and Naïve Bayes to create a model and determine criteria that will assist the BKKBN in determining a prosperous family to make it easier for field extension workers to carry out family planning counseling. From the results of data analysis, it was found that the use of naïve Bayes was 93.99% more accurate than k-NN with an accuracy of 87.28%. Based on these results, it can be concluded that the classification of prosperous families has achieved the goals of the Government's vision.*

Keywords: *Classification, K-NN, Naïve Bayes, Family Planning, Data Mining / data science*

1. PENDAHULUAN

Masalah kependudukan di Indonesia merupakan tantangan besar yang dihadapi pemerintah dari masa ke masa. Kependudukan adalah sesuatu hal yang sangat penting bagi negara yang berkaitan dengan jumlah penduduk, penyebaran, pertumbuhan, struktur, mobilitas, kualitas, dan kondisi kesejahteraan yang melibatkan ekonomi, politik, agama, budaya, sosial serta lingkungan penduduk setempat merupakan isi dari Undang-Undang (UU) Nomor 52 Tahun 2009 Tentang Perkembangan Kependudukan dan Pembangunan Keluarga, pasal 1 ayat 2.

Persoalan umum yang pasti ditemui oleh bidang kependudukan dan keluarga berencana yaitu kurang seimbangnya struktur umur penduduk dan masih tingginya angka pertumbuhan penduduk di beberapa provinsi. Penyebab utama meningkatnya pertumbuhan penduduk yang relatif tinggi yaitu masih tingginya angka kelahiran di satu pihak dan lebih banyaknya penurunan tingkat kematian di lain pihak.

Untuk mengatasi hal yang demikian, “BKKBN mengemban tugas yaitu melaksanakan tugas pemerintahan di bidang pengendalian penduduk dan penyelenggaraan keluarga berencana” sesuai Peraturan Kepala BKKBN Nomor 72/PER/B5/2011 tentang Organisasi dan Tata Kerja Badan Kependudukan dan Keluarga Berencana Nasional, dalam pasal 2[1].

Indikator yang dibuat oleh BKKBN bahwasanya ada lima kategori keluarga sejahtera tersebut, yaitu tahapan yang menjadi prioritas BKKBN adalah pra sejahtera, keluarga sejahtera I dan keluarga sejahtera II dikarenakan kategori tersebut berpotensi besar dalam peningkatan jumlah penduduk, sementara dua lainnya ialah keluarga sejahtera III dan keluarga sejahtera III-plus. Pendataan yang dilakukan oleh pendata KB (RT, Kader dan Karang Taruna) di lapangan hanya dapat mendata kategori keluarga sejahtera namun untuk pemetaan

keluarga sejahtera itu sendiri belum terklasifikasi dengan baik sehingga para penyuluh lapangan KB (PLKB) masih kesulitan untuk melakukan konseling KB.

Pada penelitian ini akan diteliti mengenai klasifikasi kategori keluarga sejahtera, Klasifikasi merupakan suatu teknik penting dalam data mining, teknik ini bisa memprediksi label kelas. Sehingga tujuan dari klasifikasi adalah kebenaran dalam memprediksi sebuah nilai [2] sedangkan Keluarga Sejahtera adalah Keluarga yang terbentuk berdasarkan perkawinan yang sah, bertaqwa kepada Tuhan Yang /maha Esa, mampu memenuhi kebutuhan hidup spiritual dan materi yang layak, memiliki hubungan yang serasi, selaras, dan seimbang antar anggota dan antar keluarga dengan masyarakat dan lingkungan[3]. Adapun data dalam penelitian ini berasal dari data PK (Keluarga Indonesia Keluarga Sejahtera) BKKBN Kota Palembang dimana data ini terdiri atas data seluruh keluarga per kecamatan di Kota Palembang.

Tahun 2005 dilakukan kajian indikator KS secara terbatas di kalangan BKKBN untuk mengakomodir berbagai saran perbaikan [4]. Hasil kajian tersebut menghasilkan beberapa perubahan indikator, adapun tahapan-tahapan indikator dari keluarga sejahtera tersebut sebagai berikut:

1. Keluarga pra sejahtera

Keluarga pra sejahtera merupakan keluarga yang belum mampu memenuhi kebutuhan dasarnya (basic need) secara minimal, seperti kebutuhan akan spiritual, sandang, pangan, papan, kesehatan dan KB.

2. Keluarga Sejahtera I

Keluarga yang sudah mampu memenuhi kebutuhan dasarnya secara minimal tapi belum mampu memenuhi kebutuhan sosial psikologinya seperti kebutuhan akan pendidikan, KB, transportasi, interaksi lingkungan tempat tinggal. Keluarga Sejahtera I adalah keluarga yang kebutuhan dasar sudah terpenuhi tapi kebutuhan sosial psikologi belum bias terpenuhi merupakan pengertain dari Keluarga Sejahtera I. Indikator Keluarga Sejahtera I adalah:

- a. Pada umumnya seluruh anggota keluarga, makan 2 kali atau lebih dalam satu hari.
- b. Saat berada di rumah, bekerja, sekolah maupun berpergian anggota keluarga mempunyai pakaian berbeda
- c. Rumah yang dihuni keluarga memiliki atap, lantai dan dinding yang baik.
- d. Dibawa ke sasaran kesehatan apabila anak sakit dan atau pasangan usia subur ingin ber KB.
- e. Pergi ke sarana pelayanan kontrasepsi apabila pasangan usia subur (PUS) ingin menggunakan KB
- f. Semua anak bersekolah khususnya anak usia 7-15 tahun di keluarga.

3. Keluarga Sejahtera II

Keluarga Sejahtera II adalah keluarga yang sudah bisa memenuhi seluruh kebutuhan dasar, kebutuhan sosial psikologis dan perkembangan keluarganya. Indikator Keluarga Sejahtera II meliputi:

- a. Anggota keluarga melakukan ibadah sesuai dengan kepercayaan dan agama masing-masing.
- b. Dalam seminggu paling kurang sekali keluarga makan daging atau ikan atau telur.
- c. Semua anggota keluarga memperoleh paling kurang 1 stel pakaian baru dalam kurun waktu setahun terakhir
- d. Untuk setiap penghuni satu rumah luas lantai rumah paling kurang 8 m².
- e. Keluarga dalam keadaan sehat sehingga bisa melakukan tugas dan fungsi masing-masing dalam kurun waktu tiga bulan terakhir.
- f. Ada satu atau lebih anggota keluarga yang bekerja untuk memperoleh penghasilan
- g. Bisa baca tulis latin bagi seluruh anggota keluarga usia 10-60 tahun.
- h. Menggunakan alat atau obat kontrasepsi bagi pasangan usia subur dengan anak dua atau lebih.

Dalam penelitian ini menggunakan metode *Fuzzy K-Nearest Neighbor*, yang akan dibandingkan dengan metode *Naive Bayes*. Metode *Fuzzy K-Nearest Neighbor* adalah pengembangan dari teori *K-Nearest Neighbor* yang digabungkan dengan teori fuzzy dalam menyampaikan pemberian label kelas pada data uji yang diprediksi [5]. Sedangkan metode *Naive Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Metode *Naive Bayes* juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi komputasi [6].

Tujuan penelitian ini adalah untuk melakukan klasifikasi dengan menggunakan algoritma *Fuzzy K-Nearest Neighbor* dan *Naive Bayes* dalam memprediksi kategori keluarga sejahtera dengan menggunakan kriteria-kriteria yang sudah ditentukan dengan mengetahui tingkatan keluarga sejahtera guna mempermudah penyuluh dalam melaksanakan konseling KB.

2. METODE PENELITIAN

2.1. Teknik Pengumpulan Data

Data primer dan sekunder merupakan data yang digunakan dalam suatu penelitian. Melakukan pengamatan langsung di lapangan dengan berpedoman pada instrumen penelitian melakukan observasi merupakan cara untuk menghasilkan data primer, sedangkan untuk mendapatkan data sekunder bisa didapatkan dari berbagai laporan dan publikasi yang relevan dengan penelitian.

Data dalam penelitian ini berasal dari data PK (Keluarga Indonesia Keluarga Sejahtera) BKKBN Kota Palembang dimana data ini terdiri atas data seluruh keluarga per kecamatan di Kota Palembang. Data tersebut merupakan data tahun 2015 sampai dengan tahun 2020 yang pada setiap tahunnya dilakukan pemuktahiran data.

2.2.1. Metode Observasi

Cara pengumpulan data dengan cara pengamatan secara langsung adalah pengertian dari metode observasi. Tidak boleh ada pengaruh terhadap objek dalam tahapan observasi [7]. Dengan maksud lain, peneliti hanya mengamati atau mengobservasi fenomena yang terjadi. Untuk memudahkan dalam observasi maka diperlukan instrumen atau daftar pengamatan. Hal ini dimaksudkan agar observasi dilakukan sesuai target penelitian dan tidak melebar ke bidang yang tidak sesuai dengan tujuan penelitian. Hal-hal yang perlu diperhatikan dalam melakukan metode observasi, yaitu :

1. Dimana akan dilakukan observasi
2. Objek apa yang akan diobservasi
3. Data-data apa yang akan dikumpulkan pada saat observasi
4. Cara observasi

2.2.2. Teknik Pengambilan Sample secara Acak Cluster

Apabila populasi tersebar dalam beberapa wilayah (*cluster*) yang memiliki kriteria yang sama maka salah satu atau beberapa wilayah dapat diambil secara acak sebagai sample. *Cluster sampling* atau sampling daerah dilakukan terhadap populasi terbesar di suatu daerah. Keuntungan metode ini adalah lebih mudah dilakukan bagi peneliti pada populasi daerah besar. Metode ini sering digunakan pada populasi yang tersebar di suatu wilayah dimana kondisi tidak dikenal semuanya[8].

2.2 Data Mining

Data mining merupakan Suatu proses penemuan informasi baru dengan cara mencari pola-pola atau aturan tertentu dari sejumlah data yang sangat besar merupakan pengertian data mining secara sederhana.[9] Untuk menggali dan menemukan pengetahuan dalam data mining terdapat enam kelompok fungsional data mining yang dapat digunakan yaitu:

1. Deskripsi memiliki fungsi bagi data yang memiliki banyak jenis dan berskala besar dalam memberi gambaran secara ringkas.
2. Estimasi berfungsi secara garis besar dalam menaksir suatu nilai yang belum diketahui.
3. Prediksi misalnya memperkirakan stok produk suatu perusahaan dalam kurun waktu satu tahun kedepan atau dapat disimpulkan menerka suatu nilai di masa mendatang.
4. Klasifikasi bertujuan untuk dapat memprediksi kelas dari suatu objek yang labelnya tidak diketahui yang memiliki proses dalam memperoleh suatu fungsi atau model yang bias membedakan kelas data atau konsep.
5. Pengelompokan bertujuan mengidentifikasi data dengan karakteristik tertentu.
6. Asosiasi sebagaimana dimaksud digunakan untuk mengidentifikasi barang-barang yang memiliki peluang dibeli konsumen bersamaan dengan produk lain dengan kata lain analisis keranjang pasar.

2.3 Metode

Metode yang dipilih pada penelitian ini berdasarkan kesesuaian data dengan metode pengklasifikasi yang paling baik yang sudah digunakan oleh beberapa peneliti sebelumnya. Berikut ini algoritma yang akan digunakan dalam penelitian ini adalah sebagai berikut :

1. Metode Fuzzy K-Nearest Neighbor

Metode yang menghasilkan hasil kesalahan yang dikategorikan sedikit merupakan metode klasifikasi yang biasa dan baik digunakan karena penentuan keputusan dalam mengklasifikasikan suatu objek sangat penting .

Metode K-Nearest Neighbor yang digabungkan dengan teori fuzzy menghasilkan metode baru yaitu Metode *Fuzzy K-Nearest Neighbor* untuk menyampaikan pemberian label kelas pada data uji yang di prediksi. Metode ini bisa memberikan prediksi secara tegas pada data yang uji berdasarkan kelas dengan memberikan nilai keanggotaan. dan tetangga terdekat. Teknik algoritma K-Nearest Neighbor juga mudah untuk

diimplementasikan. Dalam hal ini jumlah data atau biasa disebut dengan tetangga terdekat ditentukan oleh user yang dinyatakan oleh k [10]. Sebuah data bisa dimiliki oleh kelas yang berbeda dengan nilai derajat keanggotaan pada interval dan bisa dikatakan memiliki nilai keanggotaan pada setiap kelas dalam teori himpunan fuzzy. Dalam menggunakan metode Fuzzy KNearest Neighbor terdapat beberapa langkah sebagai berikut:

1. Menghitung jarak Euclidean

Rumus jarak euclidean :

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \tag{1}$$

Di mana :

$d(x_i, x_j)$ = jarak Euclidean

x_i = record ke- i

x_j = record ke- j

a_r = data ke- r

2. Diurutkan berdasarkan nilai Euclidean distance
3. K record klasifikasi terdekat ditentukan
4. Target output merupakan kelas yang mayoritas.

2. Naïve Bayes

Salah satu algoritma yang terdapat pada teknik klasifikasi merupakan pengertian algoritma Naive Bayes. Memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai Teorema Bayes Naive dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Klasifikasi Naive Bayes diasumsikan bahwa ada atau tidak adanya ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya. Bayesian classification terbukti mempunyai akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar [11]. Teorema ini digabungkan dengan Naive dimana diasumsikan kondisi antar atribut saling bebas. Adapun rumus dari teorema Bayes tersebut yaitu :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \tag{2}$$

Dimana :

X : Data dengan class yang belum diketahui.

H : Hipotesis data X merupakan suatu class spesifik.

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probability).

$P(H)$: Probabilitas hipotesis H (prior probability).

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H .

$P(X)$: Probabilitas X .

Rapid Miner adalah sebuah solusi yang digunakan untuk melakukan analisis terhadap data mining, analisis prediksi dan text mining [12]. Teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga bisa menghasilkan keputusan yang paling baik yang digunakan oleh Rapid Miner. Adapun untuk eksperimen data penelitian, peneliti menggunakan RapidMiner 9.7 untuk mengolah data dan alat bantu mengukur akurasi dari eksperimen. Langkah-langkah pada penelitian dapat dilihat pada gambar 1 dibawah ini:



Gambar 1. Tahapan Penelitian

1. Data Selection

Data yang digunakan adalah Data PK (Keluarga Indonesia Keluarga Sejahtera) BKKBN Kota Palembang dimana data ini terdiri atas data seluruh keluarga per kecamatan di Kota Palembang. . Adapun jumlah data yang diperoleh sebanyak 10.187 dibagi menjadi data training (data keluarga sejahtera) sejumlah 9.619 dan data testing sejumlah 566 data. Dalam penelitian ini data yang digunakan termasuk data dalam lingkup Pembangunan

Keluarga yang memiliki beberapa variabel data . Adapun variabel data tersebut terdiri dari 40 variable yaitu Kode Kabupaten, Kode Kecamatan, Kode Kelurahan, Kode RW, Kode RT, NIK 1 (Kepala Keluarga), Nama, Usia Kawin, Jumlah Anak, Kesertaan Ber-KB, Pakaian, Makan, Berobat, Pakaian Berbeda, Makan (Daging), Ibadah, PUS (Pasangan Usia Subur), Tabungan, Komunikasi, Hubungan Sosial, Informasi, Pengurus Kegiatan Sosial, Posyandu, BKB (Bina Keluarga Balita), BKR (Bina Keluarga Remaja), PIK-R, BKL (Bina Keluarga Lansia), UPPKS, Jenis Atap, Jenis Dinding, Jenis Lantai, Sumber Penerangan, Air Minum, Bahan bakar memasak, Toilet, Kepemilikan Bangunan, Luas Bangunan, Anggota, Kecamatan dan Bobot nilai (Y).

2. *Processing dan Cleaning*

Proses cleaning dilakukan dengan menghilangkan noise, membuang duplikasi data, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi) [13]. Setelah memperoleh data dengan jumlah 42 variabel, data tersebut dikelola untuk membuang variable-variavle yang tidak diperlukan sehingga tersisa menjadi 32 variabel yang terdiri dari Usia Kawin, Jumlah Anak, Kesertaan Ber-KB, Pakaian, Makan, Berobat, Pakaian Berbeda, Makan (Daging), Ibadah, PUS (Pasangan Usia Subur), Tabungan, Komunikasi, Hubungan Sosial, Informasi, Pengurus Kegiatan Sosial, Posyandu, BKB (Bina Keluarga Balita), BKR (Bina Keluarga Remaja), PIK-R, BKL (Bina Keluarga Lansia), UPPKS, Jenis Atap, Jenis Dinding, Jenis Lantai, Sumber Penerangan, Air Minum, Bahan bakar memasak, Toilet, Kepemilikan Bangunan, Luas Bangunan, Anggota dan Bobot nilai (Y).

Dimana variabel input adalah (X) dan varibel output adalah (Y). Adapun 32 data tersebut dibagi menjadi 4 parameter berfungsi untuk memberikan petunjuk pada hasil untuk melihat turunan (kondisi) mana yang paling dominan dalam penentuan keluarga sejahtera seperti pada tabel 1 berikut :

Tabel 1. Pembagian Parameter

Fokus	Parameter	Variabel
Pada parameter Grafis membahas kondisi keluarga dari segi keluarga berencana, yang berhubungan dalam penentuan keluarga sejahtera itu sendiri.	Grafis	X1 = Usia X2 = Jumlah Anak X3 =Kesertaan ber-KB
Parameter Keluarga membahas kondisi keluarga dari segi kebutuhan ekonomi, keharmonisan keluarga dan hubungan sosial keluarga dengan sesama masyarakat yang menjadi point utama penentuan tingkat keluarga sejahtera	Keluarga	X4 = Pakaian X5 = Makan X6 = Berobat X7 = Pakaian Berbeda X8 = Makan (Daging) X9 = Ibadah X10 = PUS (Pasangan Usia Subur) X11 = Tabungan X12 = Komunikasi X13 = Hubungan Sosial X14 = Informasi X15 = Pengurus Kegiatan Sosial X16 = Posyandu X17 = BKB (Bina Keluarga Balita)

		<p>X18 = BKR (Bina Keluarga Remaja)</p> <p>X19 = PIK-R</p> <p>X20 = BKL (Bina Keluarga Lansia),</p> <p>X21 = UPPKS</p>
<p>Terakhir parameter tempat tinggal. Parameter ini sendiri merujuk ke keadaan keluarga tersebut dari kelayakan tempat tinggal, anggota keseluruhan yang menetap dan komponen-komponen pembantu di rumah itu sendiri.</p>	Tempat Tinggal	<p>X22 = Jenis Atap</p> <p>X23 = Jenis Dinding</p> <p>X24 = Jenis Lantai</p> <p>X25 = Sumber Penerangan</p> <p>X26 = Air Minum</p> <p>X27 = Bahan bakar memasak</p> <p>X28 = Toilet</p> <p>X29 = Kepemilikan Bangunan</p> <p>X30 = Luas Bangunan</p> <p>X31 = Anggota</p>
<p>Parameter hasil adalah parameter output akhir yang menghasilkan penentuan keluarga sejahtera.</p>	Hasil	Y = Output

3. Transformation

Transformation merupakan metode yang digunakan dalam cleaning data pada tahap ini data yang telah dipisah dan dipilih lalu diubah kedalam numerical berfungsi mengubah data ke dalam format yang sesuai untuk diproses dalam data mining [14]. Sebagai penjelasan Untuk data Kesertaan ber-KB dikelompokkan menjadi 3 (tiga) kelompok. Hasil transformasi data bisa dilihat pada table 2 dibawah ini :

Tabel 2. Hasil Transformasi Data

No	Kesertaan ber-KB	Ditransformasikan menjadi nilai
1	Ya	1
2	Tidak	2
3	Tidak berlaku	3

4. Klasifikasi

Pada tahapan ini, dilakukan proses penentuan data menggunakan tools Rapidminer terhadap data penduduk yang sudah ditesting menggunakan model classifier yang sudah dibentuk.

5. Akurasi

Hasil Akurasi Hasil menunjukkan perbandingan antara dua metode yaitu metode Fuzzy K-Nearest Neighbor dan Metode Naïve Bayes.

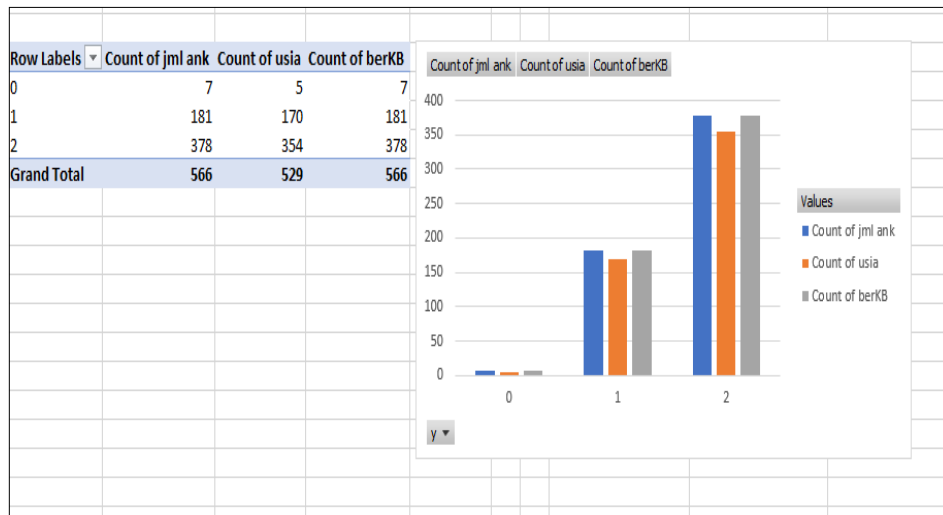
3. HASIL DAN PEMBAHASAN

3.1 Hasil Uji 4 Parameter

Untuk melihat turunan (kondisi) mana yang paling dominan dalam penentuan keluarga sejahtera, maka hasil uji dari 4 Parameter variable X adalah sebagai berikut

1. Kondisi grafis

- X1 = Usia
- X2 = Jumlah anak
- X3 = Kesertaan ber-KB

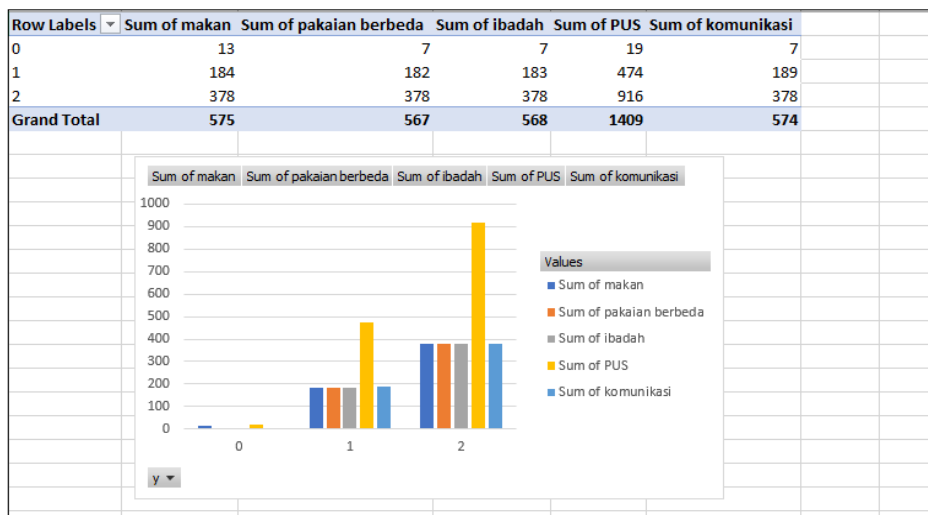


Gambar 2. Kondisi Grafis

Dari kondisi grafis, bisa dilihat pada histogram diatas variable yang paling dominan yaitu Variabel Jumlah Anak dalam penentuan keluarga sejahtera.

2. Kondisi keluarga

- X1 = Makan
- X2 = Pakaian Berbeda
- X3 = Ibadah
- X4 = PUS
- X5 = Komunikasi

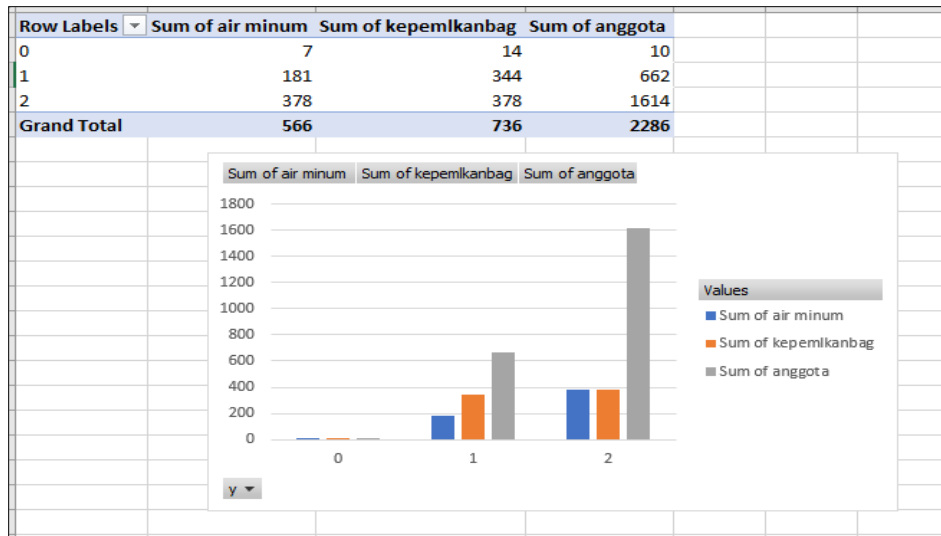


Gambar 3. Kondisi Keluarga

Dari kondisi keluarga, bisa dilihat pada histogram diatas variable yang paling dominan yaitu Variabel PUS (Pasangan Usia Subur) dalam penentuan keluarga sejahtera.

3. Kondisi tempat tinggal

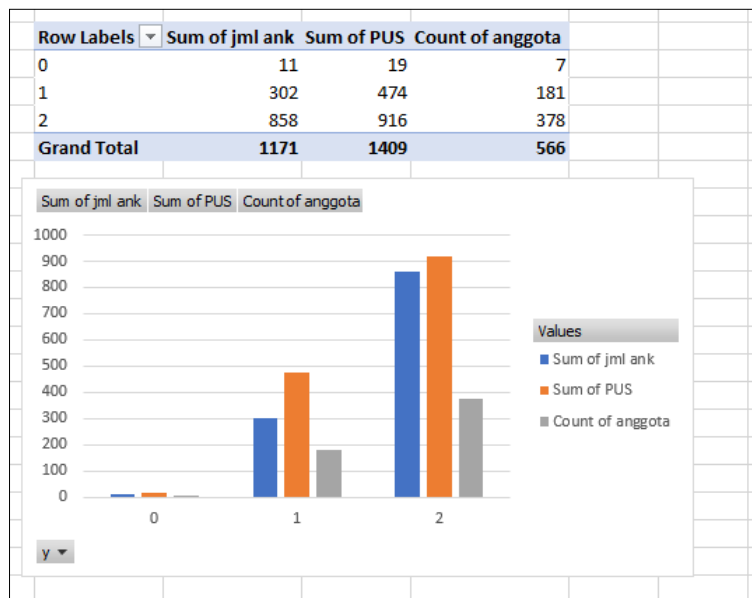
- X1 = Air Minum
- X2 = Kepemilikan Bangunan
- X3 = Jumlah Anggota Menetap di Rumah



Gambar 4. Kondisi Tempat Tinggal

Dalam parameter kondisi tempat tinggal, bisa dilihat pada histogram diatas variable yang paling dominan yaitu Variabel Jumlah Anggota yang menetap dirumah dalam penentuan keluarga sejahtera.

Selanjutnya dari setiap kondisi yang variabelnya paling dominan akan dikalkulasikan sehingga bisa diketahui variable mana yang paling dominan diantara 3 kondisi tersebut.



Gambar 5. Kondisi Perhitungan Akhir

Hasil akhir yang didapat dari diagram diatas, bahwasannya variable PUS (Pasangan Usia Subur) dalam Kondisi Keluarga memiliki nilai tertinggi, dikarenakan Pasangan Usia Subur dengan dua anak atau lebih menjadi peserta KB sangat berpengaruh dalam penentuan keluarga sejahtera. Karena jika PUS tidak ber-KB akan menambah anggota keluarga dengan bertambahnya anggota keluarga otomatis pengeluaran bertambah ditinjau dari alasan ekonomi.

3.2 Hasil Uji Pelabelan Y

Setelah dilakukan transformasi data seperti penjelasan pada tabel 2, maka label Y diuji keakuratan pelabelannya menggunakan teknik cluster. Dalam pengujian ini digunakan 100 data sampel data yang didapat dari Dinas BKKBN merupakan data real yang menghasilkan label Y (output) dengan transformasi angka 0, 1 dan 2. seperti pada tabel 3 dibawah ini :

Tabel 3. Pengujian Label Y

id	y	Cluster		
		0	1	2
1111	1	√		
1112	2	√		
1113	2	√		
.....				
1209	2	√		
1210	2	√		
Jumlah		T=68 F=13	T=12 F=4	T=2 F=1
Persentase Hitung Manual		83%	75%	67%

Setelah melakukan pengujian label Y menggunakan teknik cluster untuk meningkatkan tingkat keakuratan sebuah model yang dibangun berdasarkan dataset tertentu maka digunakan suatu teknik fold Cross Validation menggunakan 100 dataset. Berikut hasil dari perhitungan tersebut :

	true 1	true 2	true 0	class precision
pred. 1	9	0	1	90.00%
pred. 2	7	81	0	92.05%
pred. 0	0	0	2	100.00%
class recall	56.25%	100.00%	66.67%	

Gambar 6. Fold Cross validation

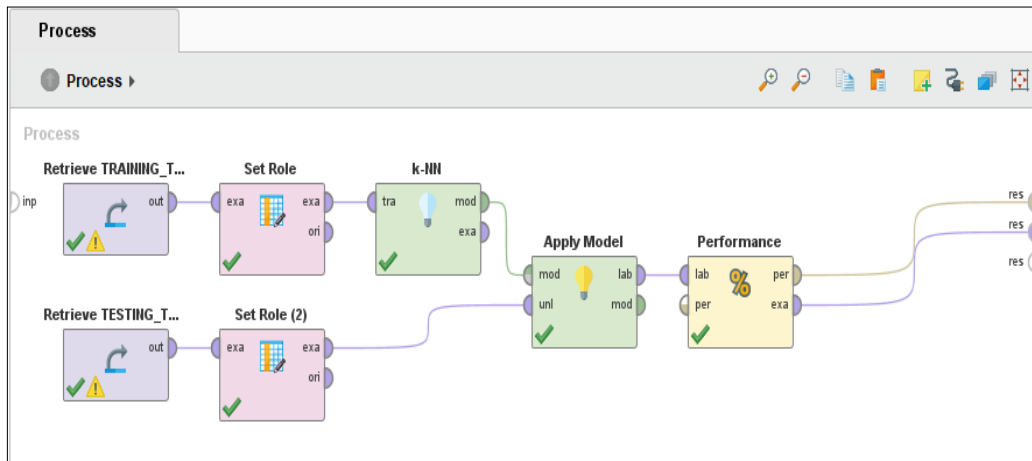
Dari hasil akhir diatas diperoleh akurasi tertinggi yaitu 91.95% dengan + 6.34% dan mendapatkan rerata micro sebesar 92 %. Adapun hasil class precision yang merupakan jumlah data yang true positive (jumlah data positif yang dikenali secara benar sebagai positif) dibagi dengan jumlah data yang dikenali sebagai positif. Dari hasil pengujian nilai precision yaitu 90.00% untuk class pred.1 , 92.05% untuk class pred.2 dan untuk class pred.0 dihasilkan nilai akurasi 100%. Dan class recall didapat 100% untuk prediksi pada class ke 2. Nilai terendah pada recall adalah pada class 1 sebesar 56.25%.

3.2. Implementasi dengan RapidMiner

Pada perhitungan ini, dilakukan pengukuran akurasi untuk membuktikan tingkat performa suatu algoritma terhadap dataset yang digunakan menggunakan alat bantu aplikasi Rapidminer.

1. Hasil pengujian Fuzzy K-Nearest Neighbor

Dalam aplikasi Rapidminer proses perhitungan performa algoritma Fuzzy K-Nearest Neighbor dapat digambarkan seperti Gambar 7 berikut:



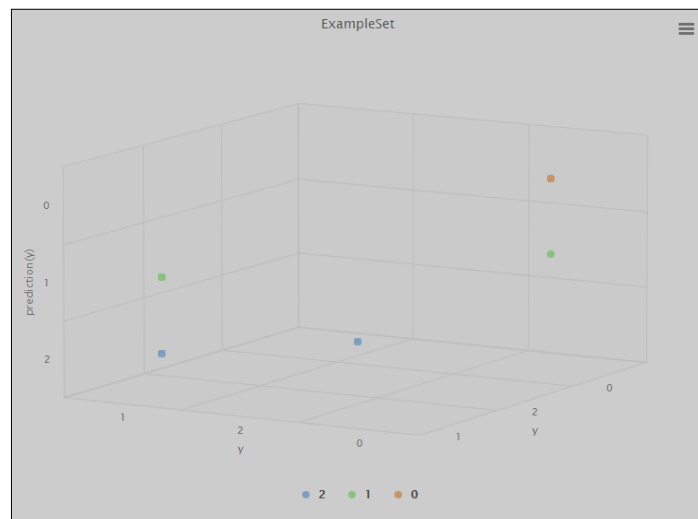
Gambar 7. Main Proses Algoritma KNN di RapidMiner

Dengan menggunakan pemodelan *Fuzzy K-Nearest Neighbor* (FK-NNC) seperti gambar di atas, dengan jumlah data training (data keluarga sejahtera) sejumlah 9.619 dan data testing sejumlah 566 data. Berikut hasil perhitungan akurasi data training menggunakan *Fuzzy K-Nearest Neighbor* (FK-NNC). Dapat diketahui tingkat akurasinya sebesar 87.28% seperti gambar 8 berikut:

accuracy: 87.28%

	true 1	true 2	true 0	class precision
pred. 1	113	1	2	97.41%
pred. 2	68	377	1	84.53%
pred. 0	0	0	4	100.00%
class recall	62.43%	99.74%	57.14%	

Gambar 8. Akurasi Data Algoritma KNN

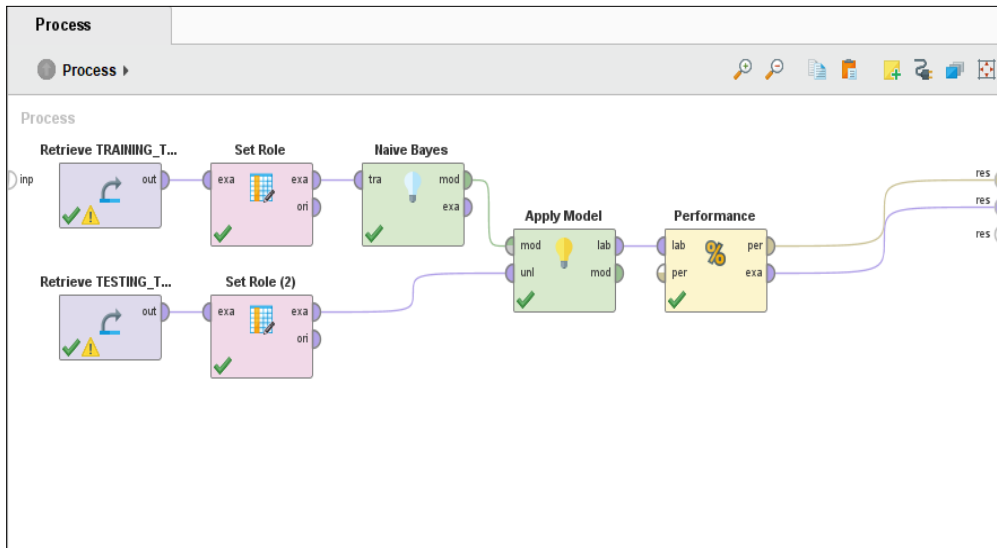


Gambar 9. Visualisasi data

Pada gambar 9 diatas, prediction (Y) yang memiliki angka nol dan berwarna orange merupakan kategori keluarga pra sejahtera, angka satu yang memiliki warna hijau yaitu kategori keluarga sejahtera 1 dan terakhir angka dua yang berwarna biru merupakan kategori keluarga sejahtera 2.

2. Hasil pengujian Naïve Bayes

Berikut Model Sub Proses di rapid Miner menggunakan agoritma Naive Bayes :

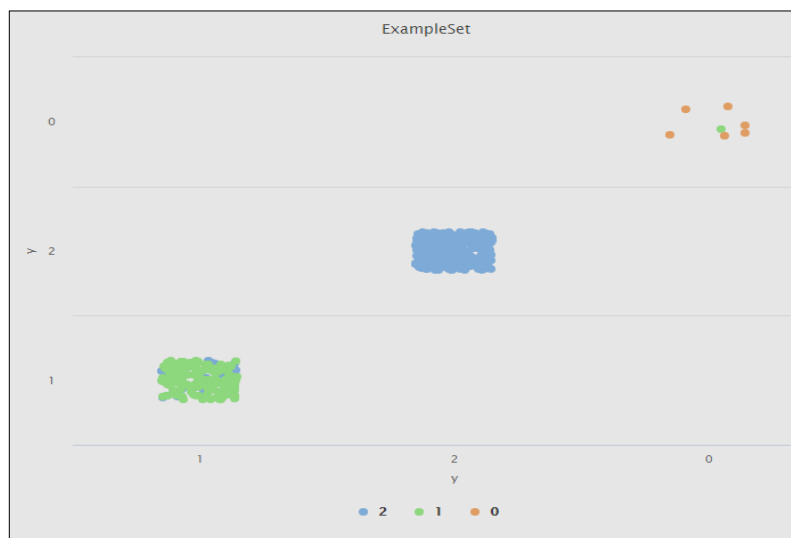


Gambar 10. Model Sub Proses Algoritma *Naive Bayes*

Selanjutnya didapatkan hasil akurasi senilai 93.99% yang bisa dilihat pada gambar 11 dibawah ini:

accuracy: 93.99%				
	true 1	true 2	true 0	class precision
pred. 1	148	0	1	99.33%
pred. 2	33	378	0	91.97%
pred. 0	0	0	6	100.00%
class recall	81.77%	100.00%	85.71%	

Gambar 11. Akurasi Data



Gambar 12. Visualisasi Data

Pada gambar 12 merupakan visualisasi data yang bertujuan untuk mendapatkan prediksi pada data yang tidak terlihat atau mengubah data dengan menerapkan model preprocessing. Dimana prediction (Y) yang memiliki angka nol dan berwarna orange merupakan kategori keluarga pra sejahtera, angka satu yang memiliki warna hijau yaitu kategori keluarga sejahtera 1 dan terakhir angka dua yang berwarna biru merupakan kategori keluarga sejahtera 2.

Berdasarkan hasil pemodelan yang telah di proses oleh tools Rapidminer selain menghasikan bentuk pola pemodelan dengan kearuratan data tools tersebut juga bisa menghasilkan data dalam bentuk visualisai. Adapun keakuratan data yang dihasilkan yaitu sebesar 93.99%.

3.3. Tingkat Akurasi Per Metode

Hasil pengujian menunjukkan bahwa pada penelitian ini diperoleh accuracy k-NN lebih tinggi dari pada accuracy Naive Bayes dengan perbandingan pada tabel 4 di bawah ini:

Tabel 4. Perbandingan Akurasi

Metode	Accuracy	Precision	Recall	Error
<i>Fuzzy K-Nearest Neighbor (FK-NNC)</i>	88.87%	93.26%	80.11%	11.13%
<i>Naive Bayes</i>	93.99%	97.10%	89.16%	6.01%

Kinerja dari Metode Naive Bayes lebih baik dari Metode *Fuzzy K-Nearest Neighbor (FK-NNC)* merujuk pada tabel 4. Bagaimanapun akurasi klasifikasi tidak dapat mencapai hasil yang sempurna dengan tidak adanya error. Hal tersebut dipengaruhi oleh banyaknya data testing dan data training yang digunakan dari tahapan preprocessing yang dilakukan.

Untuk metode Naive Bayes akurasi yang dihasilkan cukup baik, hal ini karena keunggulan dari metode Naive Bayes sendiri yaitu mampu melakukan klasifikasi meskipun memiliki data training yang sedikit untuk estimasi parameternya.

4. KESIMPULAN

Setelah menerapkan metode *Fuzzy K-Nearest Neighbor (FK-NNC)* dan Metode Naive Bayes untuk mengklasifikasikan kategori keluarga sejahtera diketahui bahwa kinerja metode Naive Bayes lebih unggul dari metode *Fuzzy K-Nearest Neighbor (FK-NNC)*. Bisa dilihat dari perhitungan dengan menggunakan Metode *Fuzzy K-Nearest Neighbor (FK-NNC)* dengan jumlah data training (data keluarga sejahtera) sejumlah 9.619 dan data testing sejumlah 566 data dihasilkan tingkat akurasinya sebesar 87.28%. Sedangkan perhitungan dengan menggunakan Metode Naive Bayes menghasilkan akurasi nilai sebesar 93.99%.

Selanjutnya kesimpulan dari sub parameter dimana variable X dibagi menjadi 4 parameter yang berfungsi untuk memberikan petunjuk pada hasil akhir untuk melihat turunan (kondisi) mana yang paling dominan dalam penentuan keluarga sejahtera. Hasil akhir yang didapat yaitu variable PUS (Pasangan Usia Subur) karena jika PUS tidak ber-KB akan menambah anggota keluarga dengan bertambahnya anggota keluarga otomatis pengeluaran bertambah ditinjau dari alasan ekonomi. Oleh karena itu Pasangan Usia Subur dengan dua anak atau lebih menjadi peserta KB sangat berpengaruh dalam penentuan keluarga sejahtera.

REFERENCES

- [1] bkkbn.go.id, "Tugas KB," *www.bkkbn.go.id*, 2020. www.bkkbn.go.id/pages/tugas-pokok-dan-fungsi (accessed Jan. 08, 2020).
- [2] S. Defiyanti, "Integrasi Metode Clustering dan Klasifikasi untuk Data Numerik," pp. 256–261, 2017.
- [3] bkkbn.go.id, "Pengertian Keluarga Sejahtera," 2020, 2020. <http://aplikasi.bkkbn.go.id/mdk/BatasanMDK.aspx> (accessed Jan. 08, 2020).
- [4] E. Sunarti, "Kajian Indikator Kesejahteraan Keluarga," Institute Pertanian Bogor, 2006.
- [5] aryono rahmad Hakim, D. Safitri, and sugito, "Klasifikasi keikutsertaan keluarga dalam program keluarga berencana (kb) di kota semarang menggunakan metode mars dan fk-nnc," vol. 5, pp. 341–349, 2016.
- [6] R. A. Saputra, A. R. Taufik, L. S. Ramdhani, R. Oktapian, and E. Marsusanti, "Sistem Pendukung Keputusan Dalam Menentukan Metode Kontrasepsi Menggunakan Algoritma Naive Bayes," *Semin. Nas. Inov. dan Tren*, pp. 106–111, 2018.

- [7] S. Herlinda, M. Said, and N. Gofar, "Metodologi Penelitian," in *Metodologi Penelitian*, L. P. U. Sriwijaya, Ed. 2015, p. 56.
- [8] S. Herlinda *et al.*, "Metodologi Penelitian," in *Metodologi Penelitian*, Lembaga Penelitian Universitas Sriwijaya, Ed. 2015, p. 51.
- [9] W. Cholil and L. Atika, "Model Data Mining Dalam Mengidentifikasi Pola Laju Pertumbuhan Antar Sektor Ekonomi di Provinsi Sumatera Selatan dan Bangka Belitung," vol. 8, no. November, pp. 103–109, 2019, doi: 10.34148/teknika.v8i2.181.
- [10] A. Satria, P. Anugerah, and C. Dewi, "Implementasi Algoritme Fuzzy K-Nearest Neighbor untuk Penentuan Lulus Tepat Waktu (Studi Kasus : Fakultas Ilmu Komputer Universitas Brawijaya)," vol. 2, no. 4, pp. 1726–1732, 2018.
- [11] Catur Suprianto dan Purnama Farida, "Deteksi Penyakit Diabetes Type II dengan naive bayes Berbasis Particle Swarm Optimization," *J. Teknol. Inf.*, vol. 9, no. 2, p., 2013.
- [12] I. W. S. Wicaksana, D. A. C, D. A. Baskoro, and L. Ambarwati, "Rapid Miner," in *Belajar Data Mining dengan Rapid Miner*, R. Sanjaya, Ed. 2013, p. 8.
- [13] T. Thi Bi Dan, S. Widya Sihwi, and R. Anggrainingsih, "Implementasi Iterative Dichotomiser 3 Pada Data Kelulusan Mahasiswa S1 Di Universitas Sebelas Maret," *J. Teknol. Inf. ITSmart*, vol. 4, no. 2, p. 84, 2016, doi: 10.20961/its.v4i2.1770.
- [14] I. N. Syahfitri, "Penerapan Data Mining untuk Menentukan Besar Pinjaman pada Koperasi Simpan Pinjam," *J. Inform. dan Sist. Inf. Univ. Ciputra*, vol. 03, no. 02, pp. 18–27, 2017.