

Klasifikasi Pemanfaat Program Beras Sejahtera (RASTRA) Berdasarkan Tingkat Kemiskinan Dengan Menggunakan Algoritma Decision Tree C4.5 Berbasis Particle Swarm Optimization

Sigit Hadi Waluyo¹⁾, Prihandoko²⁾

¹⁾ Program Magister Teknik Informatika Universitas AMIKOM Yogyakarta
Jl. Veteran, Singopuran, Kartasura, Kabupaten Sukoharjo, Jawa Tengah 57163

²⁾ Dosen Universitas Gunadarma Depok
Jl Margonda Raya No 100, Depok Jawa Barat, Indonesia

Terima Naskah : 8 September 2017
Terima Revisi : 26 September 2017

ABSTRAK

Program perlindungan sosial oleh pemerintah telah berlangsung sejak lama. Salah satunya adalah program beras sejahtera (RASTRA) yang ditujukan kepada masyarakat miskin dengan penghasilan rendah. Banyak terjadi masalah dalam prosesnya, mulai dari data masyarakat miskin yang tidak sesuai, penentuan penerima manfaat salah, dan banyak lagi lainnya. Dalam menentukan penerima manfaat program ada beberapa indikator yang dapat dijadikan sebagai pedoman, diantaranya pekerjaan, pendapatan, daya listrik lantai rumah, dinding rumah, sumber air dan kepemilikan TV. Agar proses penentuan dapat dilakukan secara akurat, cepat dan obyektif, maka diperlukan teknik untuk menentukan urutan indikator penerima manfaat program. *Decision Tree C4.5* merupakan salah satu teknik data mining untuk mengolah data pengetahuan untuk memudahkan proses penyelesaian masalah dengan menghasilkan pohon keputusan yang nantinya akan membantu pengambilan keputusan. *Decision Tree C4.5* juga akan meningkat akurasi apabila di optimasikan dengan *particle swarm optimization*. Dari hasil teknik yang dilakukan diharapkan dapat membantu proses pengambilan keputusan pihak terkait.

Kata Kunci : *Decision Tree C4.5, Particle Swarm Optimization, RASTRA.*

ABSTRACT

The government's social protection program has been going on for a long time. One of them is the Beras Sejahtera Program (RASTRA) that aimed at poor people with low income. There are many problems in the process, ranging from improperly poor data, wrong people, and many others. In determining the program beneficiaries there are several indicators that can be used as guidelines, such as kind of job, salary, electrical power, house floor, house wall, water sources and TV ownership. In order for the determination process to be accurate, fast and objective, a technique is needed to determine the sequence of program beneficiary indicators. Decision Tree C4.5 is one of the data mining techniques, to help the process of knowledge data to facilitate the problem solving process by generating decision tree to help decision making. Accuracy of Decision Tree C4.5 will increase when optimized with particle swarm optimization. From the techniques results expected to help all of stakeholders to make decision.

Keywords : *Decision Tree C4.5, Particle Swarm Optimization, RASTRA.*

PENDAHULUAN

Indonesia merupakan negara yang masih dan terus melakukan pembangunan di setiap bidang secara berkelanjutan. Hal ini tertuang pada Nawacita pemerintah yakni membangun Indonesia dari pinggiran dengan memperkuat daerah-daerah

dan desa, meningkatkan kualitas hidup manusia Indonesia, serta mewujudkan kemandirian ekonomi dengan menggerakkan sektor strategis ekonomi domestik dengan subagenda kedaulatan pangan[1]. Berbicara tentang kedaulatan pangan, pemerintah memiliki Program Perlindungan Sosial

(Social Protection) di bawah tanggung jawab Kementerian Koordinator Bidang Pembangunan Manusia dan Kebudayaan yang di dalamnya terdapat program Beras Sejahtera (RASTRA) ditujukan kepada masyarakat miskin yang memiliki pendapatan rendah. Keluarga Penerima Manfaat (KPM) berhak menerima bantuan berupa beras sebanyak 15 Kg per bulan dengan harga tebus sebesar Rp. 1.600/kg[1].

Berdasarkan data statistik kesejahteraan rakyat Kota Probolinggo yang berdasar pada hasil Susenas 2016, rumah tangga yang menerima bantuan Beras Miskin (RASKIN) mencapai 45,24 %. Jika dibandingkan dengan angka kemiskinan Kota Probolinggo tahun 2016 hanya 7,97 %, hal ini mengindikasikan bahwa keluarga penerima bantuan RASKIN tidak hanya dari keluarga miskin tetapi juga keluarga yang tidak layak menerima bantuan RASKIN. Banyak keluarga yang tergolong mampu ikut merasakan bantuan namun sebaliknya justru keluarga miskin tidak mendapatkan bantuan[2]. Sehingga diperlukan metode khusus untuk mencegah kesalahan penentuan penerima manfaat.

Decision Tree C4.5 merupakan metode dalam data mining yang menghasilkan pohon keputusan. Pohon keputusan ini nantinya membantu *decision maker* dalam mengambil keputusan secara tepat dan cepat[3]. Dan akan semakin meningkat akurasi apabila dioptimalkan menggunakan *particle swarm optimization (PSO)*[4].

Tujuan dari penelitian ini adalah membuat pohon keputusan yang tingkat akurasi optimal menggunakan *Decision Tree C4.5* berbasis *Particle Swarm Optimization (PSO)* untuk membantu dalam proses penentuan penerima manfaat bantuan RASTRA berdasarkan tingkat kemiskinan.

Agar tidak menyimpang dari rumusan masalah maka ditentukan batasan masalah dalam penelitian yang akan dilakukan. Adapun batasan masalah tersebut sebagai berikut:

1. Data keluarga penerima manfaat bantuan diambil dari kelurahan di Kota Probolinggo tahun 2016.
2. Algoritma yang digunakan adalah *Decision Tree C4.5* berbasis *Particle Swarm Optimization (PSO)*.

Pengertian Data Mining

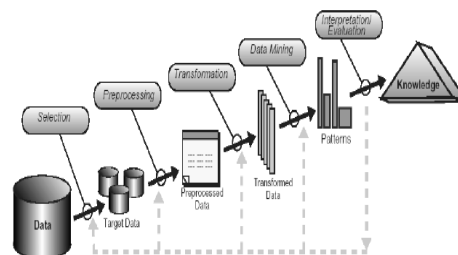
Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam data mining sangat

bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan[5]

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data. Informasi yang dihasilkan diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery in Database (KDD)*. Proses pencarian pengetahuan ini menggunakan berbagai teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksikannya. Proses pencarian bersifat iteratif dan interaktif untuk menemukan pola atau model yang benar, baru, bermanfaat, dan dimengerti. Dalam penerapannya, data mining memerlukan berbagai perangkat lunak analisis data untuk menemukan pola dan relasi data agar dapat digunakan untuk membuat prediksi dan klasifikasi dengan akurat[6].

Tahapan proses dalam penggunaan data mining yang merupakan proses *Knowledge Discovery in Database (KDD)* dapat diuraikan sebagai berikut sebagaimana ditunjukkan pada Gambar 1 [7]:

- a) Memahami domain aplikasi untuk mengetahui dan menggali pengetahuan awal serta apa sasaran pengguna.
- b) Membuat target data-set yang meliputi pemilihan data dan fokus pada sub-set data.
- c) Pembersihan dan transformasi data meliputi eliminasi derau, outliers, missing value serta pemilihan fitur dan reduksi dimensi.
- d) Penggunaan algoritma data mining yang terdiri dari asosiasi, sekuensial, klasifikasi, klusterisasi, dll.
- e) Interpretasi, evaluasi dan visualisasi pola untuk melihat apakah ada sesuatu yang baru dan menarik dan dilakukan iterasi jika diperlukan.



Gambar 1 : Proses KDD

Fungsi-fungsi yang umum diterapkan dalam data mining [8] :

- a) *Association*, adalah proses untuk menemukan aturan asosiatif antara suatu kombinasi item dalam suatu waktu.
- b) *Sequence*, hampir sama dengan *association* bedanya *sequence* diterapkan lebih dari satu periode.
- c) *Clustering*, adalah proses pengelompokan sejumlah data atau obyek ke dalam kelompok-kelompok data (klaster) sehingga setiap klaster akan berisi data yang saling mirip.
- d) *Classification*, adalah proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.
- e) *Regression*, adalah proses pemetaan data dalam suatu nilai prediksi.
- f) *Forecasting*, adalah proses pengestimasi nilai prediksi berdasarkan pola-pola di dalam sekumpulan data.
- g) *Solution*, adalah proses penemuan akar masalah dan *problem solving* dari persoalan bisnis yang dihadapi atau paling tidak sebagai informasi pendukung dalam pengambilan keputusan.

Pengertian Algoritma Decision Tree C4.5

Data merupakan representasi dunia nyata yang mewakili suatu objek seperti manusia, hewan, konsep, keadaan yang direkam dalam bentuk angka, huruf, simbol, teks, gambar, bunyi. Dengan kata lain data merupakan kenyataan yang menggambarkan suatu kejadian dan kesatuan yang nyata. Disimpulkan bahwa data merupakan bahan mentah yang akan diproses untuk menjadi sesuatu yang berguna dan bernilai [6].

Pohon Keputusan (*Decision Tree*) merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Aturan ini juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query Language (SQL)* untuk mencari *record* pada kategori tertentu. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Karena pohon keputusan memadukan antara eksplorasi data dan pemodelan, pohon keputusan ini sangat bagus

sebagai langkah awal dalam proses pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain[3].

Dalam situasi lain kemampuan untuk menjelaskan alasan pengambilan keputusan adalah sesuatu yang sangat penting. Misalnya pada perusahaan asuransi ada larangan resmi untuk mendiskriminasi berdasarkan variabel-variabel tertentu.

Perusahaan asuransi dapat mencari sendiri keadaan yang mencerminkan bahwa mereka tidak menggunakan deskriminasi yang ilegal dalam memutuskan seseorang diterima atau ditolak. Sebuah pohon keputusan adalah sebuah struktur yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Anggota himpunan hasil menjadi mirip satu dengan yang lain dengan masing-masing rangkaian pembagian.

Sebuah model pohon keputusan terdiri dari sekumpulan aturan untuk membagi sejumlah populasi yang heterogen menjadi lebih kecil, lebih homogen dengan memperhatikan pada variabel tujuannya. Sebuah pohon keputusan mungkin dibangun dengan seksama secara manual, atau dapat tumbuh secara otomatis dengan menerapkan salah satu atau beberapa algoritma pohon keputusan untuk memodelkan himpunan data yang belum terklasifikasi.

Variabel tujuan biasanya dikelompokkan dengan pasti dan model pohon keputusan lebih mengarah pada perhitungan probabilitas dari tiap-tiap *record* terhadap kategori-kategori tersebut, atau untuk mengklasifikasi *record* dengan mengelompokkannya dalam satu kelas[6].

Secara umum algoritma *Decision Tree* C4.5 untuk membangun pohon keputusan adalah sebagai berikut [6]:

1. Pilih atribut sebagai *root*
2. Buat cabang untuk masing-masing nilai
3. Bagi atribut terpilih dalam cabang
4. Ulangi proses untuk masing-masing cabang sampai semua atribut terpilih pada cabang memiliki kelas yang sama.

Untuk menghitung *entropy* diberikan rumus sebagai berikut :

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i$$

Dimana,

S = himpunan kasus

n = jumlah pastisi S

pi = proposi dari Si terhadap S

Untuk menghitung *Gain* diberikan rumus sebagai berikut :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|P_i|}{|S|} *$$

Entropy (Si)

Dimana

S = himpunan kasus

A = Atribut

n = jumlah partisi

|Si| = jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S

Untuk menghitung *gain ratio* perlu diketahui terlebih dahulu suatu term baru yang disebut dengan *Split Information* dengan rumus :

$$SplitInformation = - \sum_{t=1}^c \frac{s_t}{s} \log_2 \frac{s_t}{s}$$

Selanjutnya menghitung *Gain ratio* dengan rumus :

$$Gainratio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \quad (4)$$

Pengertian Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) merupakan teknik optimasi yang berbasis populasi yang di usulkan oleh Eberhart dan Kennedy pada tahun 1995. Metode ini terinspirasi pada perilaku sosial sekawan burung dan ikan [9]. Dalam mencari solusi yang optimal, partikel tersebut bergerak pada arah yang terbaik sebelumnya, posisi terbaik secara global. Seperti Ke-i dinyatakan $X_i = [X_{i1}, X_{i2}, \dots, X_{iD}]$ dalam ruang dimensi. Posisi terbaik sebelumnya di simpan sebagai $pbest_i = [pbest_{i1}, pbest_{i2}, \dots, pbest_{iD}]$. Indeks partikel terbaik dari semua di antara kawan group dinyatakan sebagai $gbest_d$. Kecepatan partikel dinyatakan sebagai $V_i = [V_{i1}, V_{i2}, \dots, V_{iD}]$. Modifikasi kecepatan dan posisi partikel dapat di hitung menggunakan jarak $pbest_i$, $gbest_d$ seperti ditunjukkan persamaan berikut:

$$v_{i,d} = w * v_{i,d} + c1 * R * (pbest_{i,d} - x_{i,d}) + c2 * R * (gbest_d - x_{i,d}) \quad (5)$$

$$x_{i,d} = x_{i,d} + v_{i,d} \quad (6)$$

Dimana :

$V_{i, d}$ = Kecepatan partikel ke-i pada iterasi ke-i

w = Faktor bobot inersia

c1, c2 = Konstanta akselerasi (learning rate)

R = Bilangan random (0-1)

$X_{i, d}$ = Posisi saat ini dari partikel ke-i pada iterasi ke-i

$pbest_i$ = Posisi terbaik sebelumnya dari partikel ke-i

$gbest_i$ = Partikel terbaik di antara yang terbaik pada semua partikel dalam satu kelompok atau populasi.

Pengertian Confusion Matrix dan ROC Curve

Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 4 keluaran, yaitu: *recall*, *precision*, *accuracy* dan *error rate*.

Recall adalah proporsi kasus positif yang diidentifikasi dengan benar.

$$Rumus dari Recall = d/(c+d) \quad (7)$$

Precision adalah proporsi kasus dengan hasil positif yang benar.

$$Rumus dari Precision = d/(b+d) \quad (8)$$

Accuracy adalah perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus

$$Rumus dari Accuracy = (a+c)/(a+b+c+d) \quad (9)$$

Error Rate adalah kasus yang diidentifikasi salah dengan sejumlah semua kasus

$$Rumus dari Error Rate = (b+c)/(a+b+c+d) \quad (10)$$

Keterangan :

Jika hasil prediksi negatif dan data sebenarnya negatif.

Jika hasil prediksi positif sedangkan nilai sebenarnya negatif.

Jika hasil prediksi negatif sedangkan nilai sebenarnya positif.

Jika hasil prediksi positif dan nilai sebenarnya positif.

ROC curves merupakan salah satu cara melakukan analisa terhadap model *classifier* yang telah dibuat. Penggunaan *ROC curves* adalah untuk menentukan parameter model yang diinginkan sesuai dengan karakteristik dari model *classifier* yang diinginkan. *ROC curves* dibangun menggunakan *error tipe I* dan *error tipe II*, berikut tabel dari tabel tersebut.

Tabel 1. Actual Class dan Predicted Class

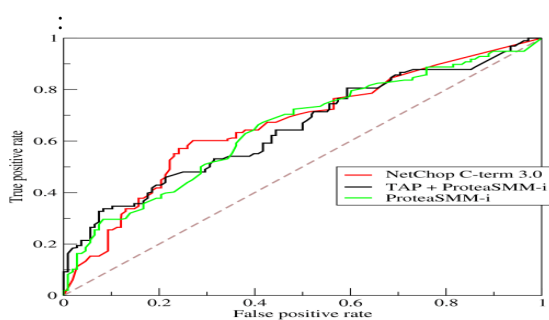
		Actual Class	
		1	0
Predicted Class	1	TP	FP
	0	FN	TN

Dari tabel 1 terdapat beberapa angka yang dapat dianalisa. Tabel terdiri dari dua buah data yaitu data kelas yang dihasilkan dari *classifier* (*Predicted Class*) dan data kelas asli yang telah diketahui (*Actual Class*). Jika kelas yang dihasilkan *classifier* sama dengan actual kelas dari data maka data termasuk dalam *True Negative* (TN) untuk kelas *positive* (1) atau *True Negative* (TN) untuk kelas *negative* (0). *ROC curves* digunakan untuk menentukan karakteristik dari *classifier*. Cara membentuknya yaitu dengan mengubah parameter-parameter yang dapat dituning dari *classifier* sehingga membuat banyak sekali *confusion table*. Dari banyak *confusion* tabel, maka kita dapat mengambil nilai dari TP dan FP dari tabel tersebut. Mengapa hanya TP dan FP? Sebenarnya bisa saja diambil perbandingan yang lain yaitu *precision* dan *recall*, tetapi pada beberapa kasus suatu *classifier* dibuat untuk menebak kelas 1 sebagai kelas yang benar sehingga kita bisa menganalisa langsung pada proses prediksi pada kelas 1 saja. Sedangkan nilai yang diplot adalah nilai *True Positive Rate* (TPR) dan *False Positive Rate* (FPR), dimana masing-masing nilai dapat dihitung dengan menggunakan persamaan sebagai berikut[10]:

$$TPR = \frac{TP}{TP+FP} \tag{11}$$

$$FPR = \frac{FP}{TP+FP} \tag{12}$$

Contoh kurva TPR vs FPR adalah sebagai berikut



Gambar 2. ROC Curve

METODE

Penelitian akan dilakukan pada data penerima bantuan Raskin tahun 2016Kelurahan Jrebeng Lor Kota Probolinggo yang beralamatkan di Jalan Sunan Ampel No.246 Kota Probolinggo.

Adapun rencana pengumpulan sumber data dengan cara sebagai berikut:

1. Sumber Data Primer
 - a. Interview
 - b. Observasi
2. Sumber Data Sekunder
 - a. Studi Dokumentasi
 - b. Studi Kepustakaan

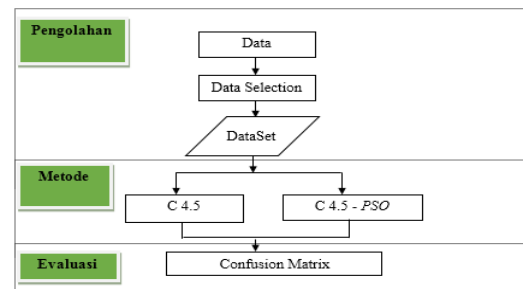
Analisa Sistem

Dalam tahapan analisa sistem ini, peneliti mengumpulkan data yang dibutuhkan guna dapat membuat pohon keputusan dari atribut yang dipakai untuk menentukan penerima manfaat bantuan. Sejauh ini sudah ada kriteria dalam menentukan siapa yang berhak mendapat bantuan. Dalam hal ini, peneliti nantinya akan menggunakan 7 kriteria.

Adapun kriteria yang dipakai untuk menentukan keluarga penerima manfaat (KPM) adalah (1) Pekerjaan, (2) Pendapatan (3) Listrik (4) Lantai (5) Dinding (6) Sumber Air (7) TV. Kemudian dari kriteria tersebut dimasukkan ke dalam excel yang kemudian dirangking berdasarkan data yang ada.

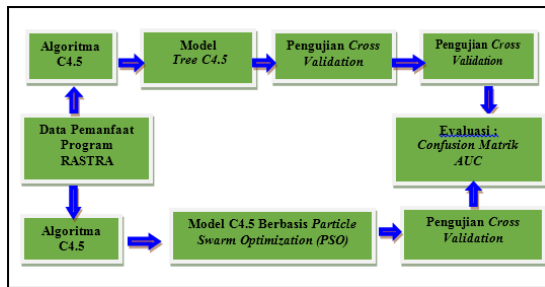
Alur Penelitian

Tahapdalam penelitian dimulai dengan pengolahan data awal, evaluasi dan berakhir dengan hasil penelitian. Perhitungan akurasi menggunakan C4.5 dan perhitungan menggunakan C4.5 berbasis *Particle Swarm Optimization* (PSO) yang kemudian dilanjutkan dengan perhitungan *Confusion Matrix* dan *ROC Curve*. Tahap ini bertujuan untuk menghasilkan pohon keputusan dan nilai akurasi yang dihasilkan. Sejauh mana hasil akurasi jika menggunakan C4.5 jikan dibandingkan dengan akurasi menggunakan C4.5 berbasis PSO. Alur tahapan metode yang akan dilakukan nantinya dapat dilihat pada gambar dibawah ini :



Gambar 3. Alur tahapan penelitian

Model penelitian ini nantinya menggunakan Algoritma C4.5 berbasis *Particle Swarm Optimization* (PSO), dengan model sebagaimana dijelaskan pada gambar 4.



Gambar 4. Model penelitian

HASIL DAN PEMBAHASAN

Hasil yang hendak dicapai dari penelitian ini adalah pohon keputusan untuk membantu dalam hal penentuan penerima manfaat RASTRA di Kelurahan Jrebeng Lor Kota Probolinggo menggunakan 7 kriteria kemiskinan. Pohon keputusan dengan nilai akurasi maksimal. Sehingga diharapkan penyaluran bantuan Beras Sejahtera (Rastra) tepat sasaran.

Dari beberapa literature yang ada, penggunaan *Algoritma Decision Tree C4.5* yang dioptimasi menggunakan *Particle Swarm Optimization* mampu menghasilkan nilai akurasi mencapai 97,67% pada kasus prediksi kelulusan mahasiswa.

SIMPULAN

Penyaluran bantuan Pemerintah kepada masyarakat melalui Program Beras Sejahtera (RASTRA) seringkali tidak tepat sasaran, banyak keluarga yang seharusnya memperoleh bantuan tidak mendapatkan bantuan, justru keluarga yang cukup ekonomi mendapat bantuan.

Masalah tersebut dapat diatasi dengan adanya pohon keputusan terhadap kriteria yang sudah ditetapkan. Analisa dan perhitungan yang tepat terhadap atribut kemiskinan yang ada dapat membantu proses penentuan keputusan

Pohon keputusan ini berupa tingkatan atribut dari kategori kemiskinan yang ada. Memberikan gambaran secara jelas prioritas atribut mana yang digunakan sebagai penentu keputusan.

DAFTAR PUSTAKA

- [1] Kementerian Koordinator Pembangunan Manusia dan Kebudayaan, Pedoman Umum Subsidi Rastra, Jakarta Pusat, 2017.
- [2] Seksi Statistik Sosial Badan Pusat Statistik Kota Probolinggo, "www.probolinggokota.bps.go.id," 22 9 2017. [Online]. Available: www.probolinggokota.bps.go.id/index.
- [3] E. Prasetyo, Data Mining Mengolah Data Menjadi Informasi menggunakan Matlab, Yogyakarta: CV. Andi Offset, 2014.
- [4] R. Wajhillah, "Optimasi Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Jantung," *swabumi*, vol. 1, 2014.
- [5] E. Frank, M. Hakk and H. Witten, Data Mining, 2011.
- [6] Kusri and E. T. Luthfi, Algoritma Data Mining, Yogyakarta: Cv. Andi Offset, 2009.
- [7] V. Kumar, J. R. J. Quinlan, Ghosh, Q. Yang, H. Motoda and G. J. N. B. L. P. Y. Z. Z. M. S. D. H. D. S. X. W. McLachan, "Top 10 Algorithms in Data Mining," vol. 14, 2008.
- [8] F. A. Hermawati, Data Mining, Yogyakarta: 2013, 2013.
- [9] H. Haryanto and L. D. Hastuti, "Metode Pohon Keputusan Menggunakan Algoritma C4.5 Untuk Pengelompokan Data Penduduk Pada Tingkat Kesejahteraan Keluarga," *Ejournal Universitas Dian Nuswantoro*, vol. 1, pp. 1-8, 2014.
- [10] W. Zhu, N. Zeng and N. Wang, "Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS Implementations," in SAS, Baltimore, 2010.