

# Penerapan Data Mining Untuk Mengklasifikasi Penerima dan Bukan Penerima Kartu Identitas Miskin (KIM) Kelurahan Sumurrejo Gunungpati dengan Metode *Naive Bayes Classifier*

Andi Bekto Rahardjo<sup>1</sup>

<sup>1,3</sup> Jurusan Teknik Informatika, FASILKOM UDINUS  
Jln. Nakula 1 No 5-11 Semarang 50131 INDONESIA

<sup>1</sup>111201106229@mhs.dinus.ac.id

**Abstract**— Results of the final public test population in 2011 recorded the number of Sumurrejo Village residents as much as 1559 households and the number of poor families (Gakin) as much as 974 households. Whereas in 2013 there were population increase as much as 1.8%, with a total population in 2013 was 5415 inhabitants from 1578 families. For gakin Village Sumurrejo in 2013 also increased as much as 5.45% with a total gakin 1027 households. The number of poor people to make the government of Semarang set of data gakin by making Poor Identity Card (KIM), which is a replacement Certificate Disadvantaged (SKTM). But there are many problems in the distribution of the KIM card, for example, in the Barusari village there are 482 households entitled to receive the distribution of KIM but until the day of card distributed, there is still no clarity. To analyze recipient of KIM card then using data mining classification techniques when they are known to reduce discrepancies in the data between the recipient and not the recipient KIM card. The method used is Naive Bayes classifier, and research design using the CRISP-DM. The research data is data poor village Sumurrejo Gunungpati Subdistrict Semarang City in 2013 were evaluated using a confusion matrix and validated by the validation split technique. The highest accuracy results obtained after research data converted is 93.06%.

**Keywords**— Identity Card Poor (KIM), data mining, Naive Bayes classifier, data mining classification, CRISP-DM

## I. PENDAHULUAN

Masyarakat miskin adalah suatu kondisi dimana fisik masyarakat yang tidak memiliki akses ke prasarana dan sarana dasar lingkungan yang memadai, dengan kualitas perumahan dan pemukiman yang jauh di bawah standart kelayakan serta mata pencaharian yang tidak menentu yang mencakup seluruh multidimensi, yaitu dimensi politik, dimensi social, dimensi lingkungan, dimensi ekonomi dan dimensi asset (P2 KP, Pedoman Umum, 2004:1).

Dari hasil final uji publik data kependudukan 2011 jumlah warga di Kelurahan Sumurrejo sebanyak 1559 KK yang terdiri dari 5319 jiwa yang terbagi dalam 6 RW dan 27 RT dengan luas wilayah 325.159 ha. Dari data tersebut tercatat jumlah warga miskin (gakin) sebanyak 974 KK. Sedangkan pada tahun 2013 terdapat kenaikan jumlah penduduk sebanyak 1,8%, dengan total penduduk tahun 2013 adalah 5415 jiwa dari 1578 KK. Untuk gakin Kelurahan Sumurrejo tahun 2013 juga mengalami peningkatan sebanyak 5,45% dengan total gakin 1027 KK. Tetapi pada tahun 2015 ini jumlah gakin Kelurahan Sumurrejo belum diketahui karena belum mencapai perhitungan tahap final, sehingga masih menggunakan acuan data pada tahun 2013.

Banyaknya warga miskin tersebut, pemerintah Kota Semarang menetapkan data gakin dengan cara membuat Kartu Identitas Miskin (KIM). Pemanfaatan KIM ini nantinya bisa

menjadi identitas bagi gakin yang akan mengakses Jamkesmaskot. Selain itu, KIM merupakan pengganti Surat Keterangan Tidak Mampu (SKTM). Tetapi pada pelaksanaannya ternyata masih ada masalah terkait pendistribusian kartu tersebut.

Dengan adanya masalah tersebut, mengakibatkan adanya ketidaksinkronan penerima kartu KIM. Masyarakat yang berhak menerima bisa jadi tidak menerima, sedangkan warga yang bukan merupakan warga miskin malah justru menerima kartu KIM. Jika penerima kartu identitas miskin (KIM) jelas terdata dengan baik, serta pembagian kartu yang terorganisasi, maka akan mengurangi ketidaksinkronan penerima kartu KIM dan juga lebih memudahkan petugas di dalam pembagian kartu KIM tersebut. Oleh karena itu dibutuhkan analisis terhadap penerima kartu KIM yang bisa dilakukan dengan teknik *data mining*.

*Data mining* diartikan sebagai ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di *database* yang sangat besar[1]. Salah satu teknik *data mining* adalah teknik klasifikasi. Teknik klasifikasi adalah teknik pembelajaran untuk prediksi suatu nilai dari target variabel kategori[2]. Karena dalam prediksi penerima kartu KIM ini memiliki dua variabel target yaitu *penerima* dan *bukan penerima* maka penelitian ini akan menggunakan *data mining* teknik klasifikasi.

*Naive bayes Classifier* merupakan salah satu metode yang digunakan dalam data mining yang didasarkan pada keputusan

*bayes*. *Naive bayes classifier* memiliki kemampuan klasifikasi seperti metode *decision tree* dan *neural network*. Metode ini dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*.

Dalam penelitian ini metode klasifikasi dengan algoritma *Naive Bayes Classifier (NBC)* akan digunakan untuk menentukan penerima dan bukan penerima kartu identitas miskin. Desain penelitian ini sendiri menggunakan model proses *CRISP-DM*.

## II. STUDI PUSTAKA

### 2.1. Penelitian Terkait

Penulis memulai penelitian dengan melakukan studi kepustakaan dari penelitian-penelitian dan sumber-sumber lain. Penelitian tersebut membahas tentang topik yang terkait dengan penelitian penulis, antara lain adalah penelitian mengenai *data mining*, dan algoritma yang digunakan penulis.

1. Dian Kartika Utami, Wisnu Ananta Kusuma, Agus Buono "Klasifikasi Metagenom dengan Metode *Naive Bayes Classifier*".
2. Sunjana "Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi *Decision Tree*".
3. S.R.Pande, Ms. S.S.Sambare, V.M.Thakre "Data Clustering Using Data Mining Techniques".
4. Dwi Widiastuti "Analisa Perbandingan Algoritma SVM, *Naive Bayes*, dan *Decision Tree* dalam Mengklasifikasikan Serangan (*Attacks*) pada Sistem Pendeteksi Intrusi".

### 2.2. Tinjauan Pustaka

#### A. Masyarakat Miskin

Masyarakat miskin adalah suatu kondisi dimana fisik masyarakat yang tidak memiliki akses ke prasarana dan sarana dasar lingkungan yang memadai, dengan kualitas perumahan dan pemukiman yang jauh di bawah standart kelayakan serta mata pencaharian yang tidak menentu yang mencakup seluruh multidimensi, yaitu dimensi politik, dimensi social, dimensi lingkungan, dimensi ekonomi dan dimensi asset (P2 KP, Pedoman Umum, 2004:1).

#### B. Data Mining

Pada jaman sekarang ini data mining sudah banyak digunakan dalam berbagai bidang termasuk informatika, dan munculnya data mining ini dikarenakan untuk mencari informasi yang berguna dan tersembunyi pada tumpukan data yang sangat besar.

Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual[3]. Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data.

Data mining merupakan irisan dari berbagai bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik, database, dan *information retrieval*. Dalam jurnal ilmiah, data mining juga dikenal dengan nama *Knowledge Discovery in Databases (KDD)*.

#### C. Cross-Industry Standard Process for Data Mining (CRISP-DM)

Cross - Industry Standard Process for Data Mining (*CRISP-DM*) dikembangkan oleh analisis dari beberapa industri pada tahun 1996. *CRISP-DM* menyediakan standar proses Data Mining sebagai strategi pemecahan masalah secara umum dari bisnis atau untuk penelitian [2]. Dalam standar ini proses Data Mining memiliki life cycle yang terbagi dalam enam fase.

1. *Business Understanding Phase*  
Pada fase ini dibutuhkan pemahaman tentang tujuan dari proyek *data mining* yang akan dilakukan. Kegiatan pada fase ini antara lain menentukan tujuan dan sasaran bisnis, menentukan kebutuhan secara detail dalam lingkup bisnis, dan menyiapkan strategi untuk mencapai tujuan.
2. *Data Understanding Phase*  
Fase pengumpulan data awal, dan menganalisis data untuk mengenal data yang akan digunakan. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal, serta mengevaluasi kualitas data.
3. *Data Preparation Phase*  
Menyiapkan kumpulan data yang akan digunakan untuk kebutuhan fase berikutnya. Fase ini merupakan pekerjaan berat dan perlu dilakukan secara intensif. Siapkan variabel yang akan ditransformasikan pada database baru dan lakukan perubahan variabel jika dibutuhkan.
4. *Modeling Phase*  
Memilih dan mengaplikasikan teknik pemodelan yang sesuai. Kalibrasi aturan model untuk mengoptimalkan hasil.
5. *Evaluation Phase*  
Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama dan untuk mendapatkan kualitas dan efektifitas sebelum disebarkan.
6. *Deployment Phase*  
Pembuatan laporan dari model yang dihasilkan dan penerapan proses Data Mining.

#### D. Naive Bayes Classifier (NBC)

*NBC* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu

class. NBC didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa *decision tree* dan *neural network*. Naïve Bayes memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes

Teorema Bayes memiliki rumusan umum sebagai berikut:

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

Dimana :

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 kondisi pada hipotesis H

H

P(X)=probabilitas dari X

### E. Confusion Matrix

*Confusion matrix* adalah tool yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah [10]. Sebuah matrix dari prediksi akan dibandingkan dengan kelas yang asli dari inputan, dengan kata lain berisi informasi nilai aktual dan prediksi pada klasifikasi.

Tabel 2.1 Tabel *confusion matrix* 2 kelas

Classification	Predicted class	
	Class = Yes	Class = No
Class = Yes	a ( <i>true positive-TP</i> )	b ( <i>true negative-TN</i> )
Class = No	c ( <i>false positive-FP</i> )	d ( <i>false negative-FN</i> )

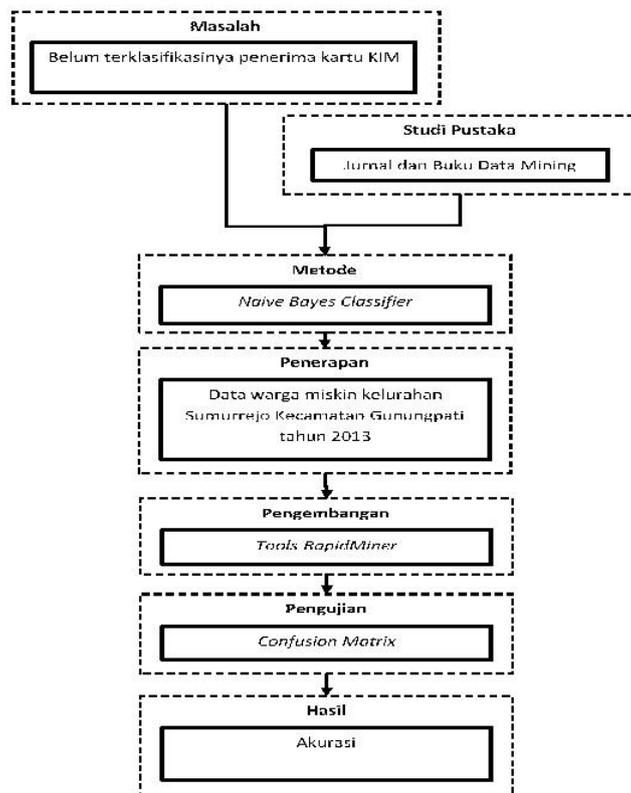
### F. Cross Validation

Cross Validation adalah teknik validasi dengan membagi data secara acak ke dalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi. Dalam Cross Validation, jumlah tetap lipatan atau partisi dari data ditentukan sendiri. Cara standar untuk memprediksi error rate dari teknik pembelajaran dari sebuah sampel data tetap adalah dengan menggunakan tenfold cross validation.

Dengan *tenfold cross validation*, data akan dibagi secara acak menjadi 10 bagian, dimana *class* diwakili (kurang lebih) proporsi yang sama seperti pada dataset yang penuh. Setiap bagian mendapatkan gilirannya dan skema pembelajaran dilatih pada sisa sembilan persepuluh; kemudian *error rate* dihitung pada *holdout set*. Dengan demikian, prosedur pembelajaran dilaksanakan sebanyak 10 kali di *training set* yang berbeda (setiap set memiliki banyak kesamaan dengan

yang lain). Akhirnya, 10 estimasi error dirata-rata untuk menghasilkan perkiraan kesalahan keseluruhan.

## 2.3. Kerangka Pemikiran



Gambar 2.1 Kerangka Pemikiran

## III METODE PENELITIAN

### 3.1 Jenis dan Sumber Data

#### 3.1.1 Data Primer

Data Primer yang digunakan dalam penelitian ini merupakan data yang diperoleh langsung dari sumber data. Data tersebut adalah *database* warga miskin Kota Semarang Kelurahan Sumurrejo Kecamatan Gunungpati tahun 2013 yang didapat dari Kelurahan Sumurrejo Kota Semarang.

#### 3.1.2 Data Sekunder

Untuk data sekunder yang berfungsi untuk membantu penyusunan tugas akhir dalam penelitian ini antara lain adalah *e-book*, jurnal, buku, beserta kumpulan materi yang membahas tentang *data mining*.

### 3.2 Metode Pengumpulan Data

#### 3.2.1 Wawancara

Penulis melakukan wawancara dengan menanyakan langsung kepada petugas Kelurahan Sumurrejo mengenai masyarakat miskin di kelurahan tersebut, bagaimana penyaluran Kartu Identitas Miskin, apakah sudah tepat sasaran atau belum, apa saja masalah yang timbul tentang adanya Kartu Identitas Miskin tersebut.

#### 3.2.2 Dokumentasi

Dokumentasi penelitian didapat setelah wawancara. Dokumentasi yang didapat antara lain data warga miskin, syarat menerima KIM, contoh kartu KIM.

#### 3.2.3 Studi Pustaka

Studi pustaka yang digunakan penulis sebagai salah satu metode pengumpulan data adalah dengan menggunakan buku, jurnal, artikel, serta kumpulan materi yang berhubungan dengan KIM, *data mining*, dan *Naive Bayes Classifier*.

### 3.3 Desain Penelitian

Standar proses *data mining* model CRISP-DM (*Cross Industry Standard Process*) yang digunakan dalam penelitian ini memiliki beberapa langkah-langkah, yaitu:

#### 1. Business Understanding

Kartu Identitas Miskin (KIM) merupakan salah satu kartu dari pemerintah Kota Semarang yang ditujukan kepada warga miskin Kota Semarang untuk lebih memudahkan di dalam menerima bantuan. Salah satu contohnya adalah warga miskin desa Karanggeneng, Gunungpati yang berjumlah 68 kepala keluarga pada tahun 2013. Tetapi untuk tahun 2014 ini masih menggunakan data pada tahun lalu, padahal bisa saja terjadi perubahan ekonomi pada warga miskin tersebut. Oleh karena itu perlu dilakukan pengklasifikasian ulang data warga miskin yang berhak menerima Kartu Identitas Miskin (KIM) untuk meminimalisir data yang tidak sinkron dengan kondisi di lapangan.

#### 2. Data Understanding

Data dalam penelitian ini diperoleh dari Kelurahan Sumurrejo Kota Semarang. Data yang didapat adalah *database* warga miskin Kecamatan Gunungpati Kota Semarang tahun 2013. Data tersebut berisikan atribut antara lain no.KK, NIK, nama, ttl, status perkawinan, jenis kelamin, alamat, pekerjaan, luas bangunan per m<sup>2</sup>, jenis lantai bangunan, ketersediaan wc, sumber penerangan, sumber air minum, bahan bakar, frekuensi makan daging, rata-rata pakaian baru dalam setahun, frekuensi makan dalam sehari, kesanggupan

biaya pengobatan, penghasilan KK, pendidikan terakhir KK, dan jumlah harta yang dimiliki.

Tabel 3.1 Atribut Data Penelitian

<b>No. KK</b>	atribut yang menginformasikan nomor kartu keluarga dari warga miskin
<b>NIK</b>	atribut yang menginformasikan nomor induk kependudukan dari warga miskin
<b>Nama</b>	atribut yang menginformasikan nama lengkap warga miskin
<b>TTL</b>	atribut yang menginformasikan tempat dan tanggal lahir dari warga miskin
<b>Status Perkawinan</b>	atribut ini menginformasikan status perkawinan dari warga miskin, dimana S (Sudah Menikah), B (Belum Menikah), P (Pernah Menikah)
<b>Jenis Kelamin</b>	atribut ini menginformasikan jenis kelamin warga miskin, yaitu Lk untuk laki-laki dan Pr untuk perempuan
<b>Alamat</b>	atribut ini menginformasikan alamat dari warga miskin
<b>Pekerjaan</b>	atribut ini menginformasikan pekerjaan dari warga miskin dari buruh, petani, kuli bangunan, serabutan, dll.
<b>Luas Bangunan per m<sup>2</sup></b>	atribut ini menginformasikan luas bangunan warga miskin yang ditempati

<b>Jenis Lantai Bangunan</b>	atribut ini menginformasikan jenis lantai bangunan warga miskin, masih tanah atau sudah keramik
<b>Ketersediaan WC</b>	atribut ini menginformasikan ketersediaan WC warga miskin, sudah memiliki sendiri atau menggunakan fasilitas lain
<b>Sumber Penerangan</b>	atribut ini menginformasikan sumber penerangan warga miskin yang ada yaitu listrik dan non-listrik
<b>Sumber Air Minum</b>	atribut ini menginformasikan sumber air minum yang digunakan seperti air pam dan air sumur
<b>Bahan Bakar</b>	atribut ini menginformasikan bahan bakar yang digunakan warga miskin untuk memasak seperti gas, kayu, dsb
<b>Rata-rata pakaian baru dalam setahun</b>	atribut ini menginformasikan jumlah pakaian baru yang dibeli warga miskin dalam setahun
<b>Frekuensi makan daging</b>	atribut ini menginformasikan frekuensi makan daging warga miskin
<b>Frekuensi makan dalam sehari</b>	atribut ini menginformasikan frekuensi makan warga miskin dalam sehari
<b>Kesanggupan biaya pengobatan</b>	atribut ini menginformasikan kesanggupan warga miskin dalam membayar biaya pengobatan
<b>Penghasilan KK</b>	atribut ini menginformasikan jumlah penghasilan KK warga miskin

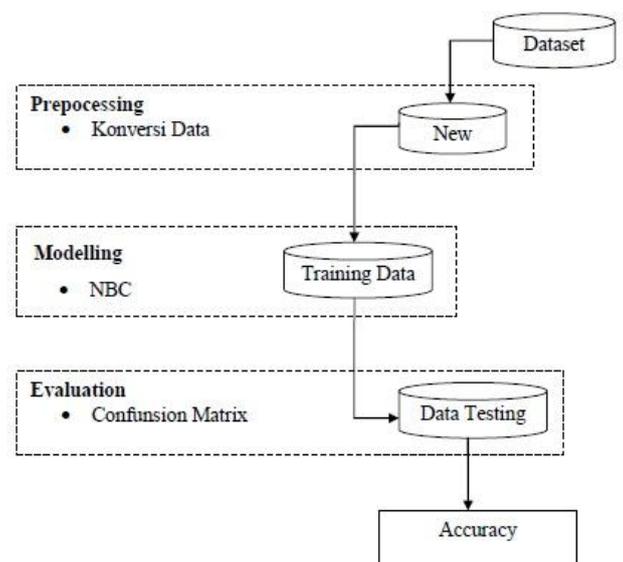
<b>Pendidikan Terakhir KK</b>	atribut ini menginformasikan pendidikan terakhir yang ditempuh oleh KK warga miskin
<b>Jumlah harta yang dimiliki</b>	atribut ini menginformasikan jumlah harta yang dimiliki warga miskin baik tunai maupun non-tunai
<b>Keterangan</b>	atribut ini menginformasikan keterangan tentang warga penerima dan bukan penerima

### 3. Data Preparation

Dari 22 atribut pada *data understanding* hanya akan digunakan 15 atribut untuk data penelitian, yaitu No.KK, pekerjaan, luas bangunan per m<sup>2</sup>, jenis lantai bangunan, ketersediaan wc, sumber penerangan, sumber air minum, bahan bakar, frekuensi makan daging, rata-rata pakaian baru dalam setahun, frekuensi makan dalam sehari, kesanggupan biaya pengobatan, penghasilan kk, jumlah harta yang dimiliki, keterangan penerima atau bukan penerima sebagai label target.

### 4. Modelling

Metode yang akan digunakan dalam penelitian ini adalah metode NBC. Kemudian pengukuran akurasi dalam penelitian ini akan menggunakan framework RapidMiner Ver. 6.0.008.



Gambar 3.1 Model Penelitian yang Diusulkan

5. Validasi dan evaluasi

Dalam tahap ini akan dilakukan validasi serta pengukuran keakuratan hasil yang dicapai oleh model menggunakan beberapa teknik yang terdapat dalam *framework* Rapid Miner Ver. 6.0.008 yaitu *Confusion Matrix* untuk pengukuran tingkat akurasi model, dan *Split Validation* untuk validasi.

IV HASIL PENELITIAN DAN PEMBAHASAN

4.1 Pengolahan Data

Data yang digunakan pada penelitian ini adalah data warga miskin Kota Semarang Kelurahan Sumurrejo Kecamatan Gunungpati tahun 2013. Data yang diperoleh dari Kelurahan masih berupa raw data. Data murni dari Kelurahan yang berisi atribut No.KK, NIK, nama, TTL, status perkawinan, jenis kelamin, alamat, pekerjaan, luas bangunan per m<sup>2</sup>, jenis lantai bangunan, ketersediaan wc, sumber penerangan, sumber air minum, bahan bakar, frekuensi makan daging, rata-rata pakaian baru dalam setahun, frekuensi makan dalam sehari, kesanggupan biaya pengobatan, penghasilan kk, pendidikan terakhir kk, jumlah harta yang dimiliki, keterangan penerima atau bukan penerima. Lalu untuk memudahkan proses *mining*, dilakukanlah tahap *preprocessing* yang berisi pemilihan atribut dan juga konversi data. Tahap pemilihan atribut dilakukan dengan cara memilih atribut apa saja yang akan digunakan sebagai *data training*. Dalam penelitian ini atribut yang digunakan berjumlah lima belas (15) antara lain No.KK, pekerjaan, luas bangunan per m<sup>2</sup>, jenis lantai bangunan, ketersediaan wc, sumber penerangan, sumber air minum, bahan bakar, frekuensi makan daging, rata-rata pakaian baru dalam setahun, frekuensi makan dalam sehari, kesanggupan biaya pengobatan, penghasilan kk, jumlah harta yang dimiliki, keterangan penerima atau bukan penerima, sehingga ada tujuh atribut yang tidak dipakai yaitu NIK, nama, TTL, status perkawinan, jenis kelamin, alamat, pendidikan terakhir KK. Sedangkan untuk data konversi akan dilakukan perubahan terhadap data dalam bentuk kategorikal untuk memudahkan proses *mining*, atribut yang diubah antara lain tempat lahir, umur, alamat, pekerjaan, dan keterangan.

Tabel 4. 1 Tabel *Sample Data* Setelah Konversi

NO.KK	FEKES REJAS N	LICAS BANGUNA N/M	JENIS LANTAI BANGUNA N	KETERSED IAAN WC	SUMBER PENERANG AN	SUMBER RAIR MINUM	BAHAN N BAKAR	FREKVEN SI MAKAN DAGING	FREKVEN SI MAKAN DALAM SEHARI	RATA- RATA PAKAIAN BARU DALAM SEHARI	KESANGGUP AN BAYAR PENGOBATA N	PENG HASIL LAN KK	JUMLAH HARTA YANG DIMILIKI	KETERANGA N
3374120409140000	1	3	4	1	3	3	3	2	2	3	2	3	1	2
3374120310060015	2	2	4	1	2	2	3	2	2	2	1	2	1	2
3374120310060017	3	2	4	1	3	3	3	2	2	1	1	3	1	2
3374120310060075	3	2	4	1	3	3	3	2	2	2	1	3	1	2
3374031210060077	3	2	4	1	2	2	3	2	2	2	1	3	1	2
3374121104140000	3	1	4	1	3	3	3	2	2	2	1	3	1	2
3374121104070000	3	1	2	2	3	1	2	1	1	2	1	2	1	1
3374121104090000	1	2	4	1	3	3	3	2	2	3	1	3	2	2
3374120310060000	4	1	4	1	3	2	3	2	1	2	1	2	1	2
3374120310060004	3	2	4	1	3	3	3	2	2	1	1	2	1	2
3374120301080004	3	1	2	2	3	1	1	1	1	1	1	2	1	1
3374120310060090	4	2	4	1	3	3	3	2	2	3	1	3	1	2
3374120310060098	3	1	4	1	3	2	3	2	2	2	1	3	2	2
Da	Da	Da	Da	Da	Da	Da	Da	Da	Da	Da	Da	Da	Da	Da

4.2 Pembahasan Hasil Percobaan

Percobaan pada penelitian ini menggunakan *tools* RapidMiner Ver 6.0.008. Algoritma yang digunakan adalah

*Naive Bayes Classifier* dan validasi yang dipakai adalah *split validation*, untuk mengukur performansinya menggunakan *confusion matrix*.

Hasil dari tiga percobaan yang sudah dilakukan menggunakan model *naive bayes*, dapat dilihat bahwa data yang sudah melewati proses pemilihan data, pemilihan atribut dan juga konversi data dari data yang berjumlah 240 *record* dan 15 atribut menghasilkan tingkat akurasi tertinggi yaitu 93,06% yang sudah termasuk dalam rentang akurasi *excellent classification*.

Untuk membuktikan hasil klasifikasi kartu KIM warga miskin sebagai penerima, bisa dilakukan dengan perhitungan manual *naive bayes classifier* sebagai berikut :

- Menghitung jumlah kelas dari klasifikasi yang terbentuk  
 C1 = Penerima  
 C2 = Bukan Penerima  
 Perhitungan :  
 $P(X|Ci) = P(\text{penerima}) = 68/240 = 0,283$   
 $P(X|Ci) = P(\text{bukan penerima}) = 172/240 = 0,717$
- Menghitung jumlah kasus yang sama dari kelas X (pekerjaan=karyawan swasta, luas bangunan=0-50 m<sup>2</sup>, jenis lantai bangunan=>50% tanah, ketersediaan wc=jamban umum, sumber penerangan=450 watt, sumber air minum=air tidak dilindungi, bahan bakar=kompur minyak, frekuensi makan daging=tidak menentu, frekuensi makan dalam sehari=2x, rata-rata pakaian baru dalam setahun=2 stel, kesanggupan biaya=jamkesmas, penghasilan kk=>500 ribu- 1 juta, jumlah harta yang dimiliki=<=10 juta).

Perhitungan :  
 $P(\text{pekerjaan}=\text{''karyawan swasta''} | \text{penerima}) = 45/68 = 0,662$   
 $P(\text{pekerjaan}=\text{''karyawan swasta''} | \text{bukan penerima}) = 91/172 = 0,530$   
 $P(\text{luas bangunan}=\text{''0-50 m}^2\text{''} | \text{penerima}) = 61/68 = 0,897$   
 $P(\text{luas bangunan}=\text{''0-50 m}^2\text{''} | \text{bukan penerima}) = 121/172 = 0,703$   
 $P(\text{jenis lantai bangunan}=\text{''>50% tanah''} | \text{penerima}) = 25/68 = 0,368$   
 $P(\text{jenis lantai bangunan}=\text{''>50% tanah''} | \text{bukan penerima}) = 0/172 = 0$   
 $P(\text{lantai ketersediaan wc}=\text{''jamban umum''} | \text{penerima}) = 35/68 = 0,515$   
 $P(\text{lantai ketersediaan wc}=\text{''jamban umum''} | \text{bukan penerima}) = 4/172 = 0,023$   
 $P(\text{sumber penerangan}=\text{''450 watt''} | \text{penerima}) = 61/68 = 0,897$   
 $P(\text{sumber penerangan}=\text{''450 watt''} | \text{bukan penerima}) = 50/172 = 0,291$   
 $P(\text{sumber air minum}=\text{''air tidak dilindungi''} | \text{penerima}) = 65/68 = 0,956$

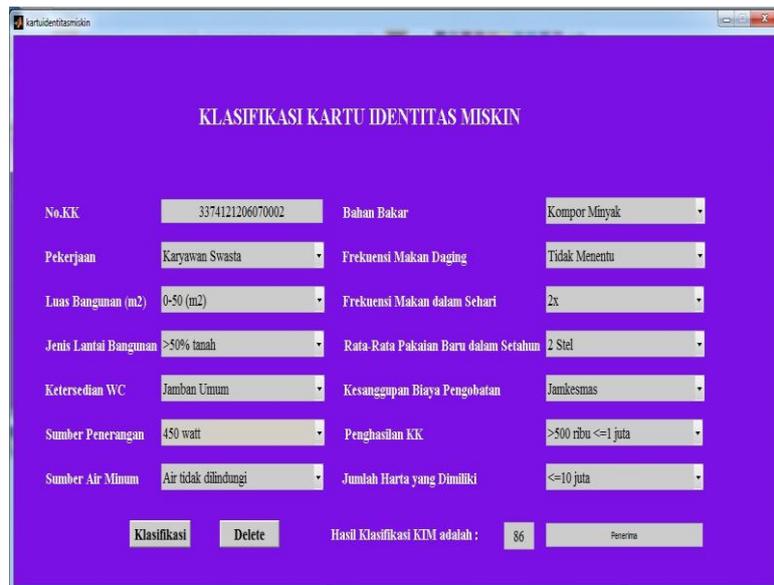
$P(\text{sumber air minum}=\text{"air tidak dilindungi"} \mid \text{bukan penerima}) = 30/172 = 0,174$   
 $P(\text{bahan bakar}=\text{"kompore minyak"} \mid \text{penerima}) = 1/68 = 0,15$   
 $P(\text{bahan bakar}=\text{"kompore minyak"} \mid \text{bukan penerima}) = 1/172 = 0,006$   
 $P(\text{frekuensi makan daging}=\text{"tidak menentu"} \mid \text{penerima}) = 68/68 = 1$   
 $P(\text{frekuensi makan daging}=\text{"tidak menentu"} \mid \text{bukan penerima}) = 138/172 = 0,802$   
 $P(\text{frekuensi makan dalam sehari}=\text{"2x"} \mid \text{penerima}) = 68/68 = 1$   
 $P(\text{frekuensi makan dalam sehari}=\text{"2x"} \mid \text{bukan penerima}) = 24/172 = 0,140$

$P(\text{rata-rata pakaian baru dalam setahun}=\text{"2 stel"} \mid \text{penerima}) = 29/68 = 0,426$   
 $P(\text{rata-rata pakaian baru dalam setahun}=\text{"2 stel"} \mid \text{bukan penerima}) = 28/172 = 0,13$   
 $P(\text{keanggupan biaya}=\text{"jamkesmas"} \mid \text{penerima}) = 68/68 = 1$   
 $P(\text{keanggupan biaya}=\text{"jamkesmas"} \mid \text{bukan penerima}) = 171/172 = 0,994$   
 $P(\text{penghasilan kk}=\text{">500 ribu - 1 juta"} \mid \text{penerima}) = 67/68 = 0,985$   
 $P(\text{penghasilan kk}=\text{">500 ribu - 1 juta"} \mid \text{bukan penerima}) = 37/172 = 0,215$   
 $P(\text{jumlah harta yang dimiliki}=\text{"<=10 juta"} \mid \text{penerima}) = 68/68 = 1$   
 $P(\text{jumlah harta yang dimiliki}=\text{"<=10 juta"} \mid \text{bukan penerima}) = 59/172 = 0,343$

3. Mengkalikan semua hasil dari atribut  
 $P(X \mid \text{penerima}) = 0,662 \times 0,897 \times 0,368 \times 0,515 \times 0,897 \times 0,956 \times 0,15 \times 1 \times 0,426 \times 1 \times 0,985 \times 1 = 0,0061$   
 $P(X \mid \text{bukan penerima}) = 0,530 \times 0,703 \times 0,023 \times 0,291 \times 0,174 \times 0,006 \times 0,802 \times 0,140 \times 0,13 \times 0,215 \times 0,343 = 0$   
 $P(X \mid \text{penerima}) \times P(\text{penerima}) = 0,0061 \times 0,283 = 0,002$   
 $P(X \mid \text{bukan penerima}) \times P(\text{bukan penerima}) = 0 \times 0,530 = 0$

4. Bandingkan hasil kelas :  
 $P(X \mid \text{penerima}) \times P(\text{penerima})$   
 $P(X \mid \text{bukan penerima}) \times P(\text{penerima})$   
 Kesimpulan akhir : dengan data warga miskin seperti diatas maka warga tersebut digolongkan dalam klasifikasi penerima kartu KIM karena hasil probabilitas  $P(X \mid \text{penerima}) \times P(\text{penerima})$  lebih besar dari yang bukan penerima

#### 4.3 Implementasi pada Matlab



Gambar 4. 1 Tampilan pada GUI Matlab

Pada tampilan GUI (*Graphical User Interface*) gambar 4.1 diatas, aplikasi dijalankan dengan memasukkan inputan antara lain adalah no kk, pekerjaan, luas bangunan, jenis lantai bangunan, ketersediaan wc, sumber penerangan, sumber air minum, bahan bakar, frekuensi makan daging, frekuensi makan dalam sehari, rata-rata pakaian dalam setahun, kesanggupan biaya pengobatan, penghasilan kk, dan jumlah harta yang dimiliki. *Button* klasifikasi digunakan untuk melihat hasil klasifikasi KIM yaitu penerima atau bukan penerima dengan hasil skor sesuai dengan inputan. Sedangkan *button delete* digunakan untuk menghapus inputan, skor, dan hasil klasifikasi apabila ingin melakukan inputan kembali.

## V KESIMPULAN DAN SARAN

### a. Kesimpulan

Berdasarkan penelitian dilakukan peneliti, maka dapat disimpulkan beberapa hal sebagai berikut :

1. Peneliti memperoleh Data Warga Miskin Kota Semarang Kelurahan Sumurejo Kecamatan Gunungpati yang berupa data printout dari kelurahan, setelah data print out tersebut diolah dengan metode data mining naive bayes, Peneliti dapat menyimpulkan bahwa dari 240 data warga miskin Desa Karanggeneng Kelurahan Sumurrejo di peroleh 63 warga miskin Penerima kartu KIM dan sisanya 177 warga Bukan Penerima Kartu KIM.
2. Peneliti telah membuat aplikasi yang bisa digunakan untuk menentukan penerima dan bukan penerima Kartu Identitas Miskin (KIM) sesuai dengan atribut yang sudah dipilih antara lain No.KK, pekerjaan, luas bangunan per m<sup>2</sup>, jenis lantai bangunan, ketersediaan wc, sumber

penerangan, sumber air minum, bahan bakar, frekuensi makan daging, rata-rata pakaian baru dalam setahun, frekuensi makan dalam sehari, kesanggupan biaya pengobatan, penghasilan kk, dan jumlah harta yang dimiliki.

3. Tingkat akurasi data setelah dikonversi yang didapatkan setelah dilakukan evaluasi dengan *confusion matrix* adalah 93,06% yang merupakan akurasi tertinggi dan termasuk dalam rentang akurasi *excellent classification*..

b. Saran

Berdasarkan penelitian yang dilakukan dan agar penelitian ini dapat berkembang, peneliti memberikan saran sebagai berikut :

1. Melakukan pengembangan penelitian dengan metode klasifikasi *data mining* selain *naïve bayes classifier*.

2. Melakukan penambahan *record* pada penelitian selanjutnya untuk menambah tingkat akurasi.
3. Diharapkan dengan adanya aplikasi yang sudah dibuat oleh peneliti bisa mempermudah pegawai Kelurahan Sumurrejo di dalam menentukan penerima KIM sehingga bisa tepat sasaran.

UCAPAN TERIMA KASIH

Penulis mengucapkan teima kasih kepada Universitas Dian Nuswantoro, Rektor UDINUS, Dekan Fakultas Ilmu Komputer, Kaprodi Teknik Informatika-S1, Dosen pembimbing, Dosen-dosen pengampu kuliah di Fakultas Ilmu Komputer , serta teman-teman dan sahabat yang selama ini telah mendampingi penulis selama kuliah di Universitas Dian Nuswantoro.

REFERENSI

- [1] Yudho Giri Sucahyo (2003). IlmuKomputer.Com.
- [2] Kusriani and E.T Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Offset,2009
- [3] H. Santoso, “Analisis dan Prediksi pada Perilaku Mahasiswa Diploma untuk Melanjutkan Studi ke Jenjang Sarjana Menggunakan Teknik Decision Tree dan Support Vektor Machine” , Agust 2012
- [4] Dian Kartika Utami, Wisnu Ananta Kusuma and Agus Buono, “Klasifikasi Metagenom dengan Metode Naive Bayes Classifier”, 2014
- [5] Sunjana, “Aplikasi Mining Data Mahasiswa dengan Metode Klasifikasi Decision Tree”, 2010
- [6] Larose D. T, *Discovering Knowledge in Database*. New Jersey: John Wiley and Sons Inc, 2005
- [7] S.R Pande, Ms. S.S.Sambare and V.M. Thakre, “Data Clustering using Data Mining Techniques”, 2012
- [8] Dwi Widiastuti, “Analisa Perbandingan Algoritma SVM, Naive Bayes, dan Decision Tree dalam Mengklasifikasikan Serangan (Attacks) pada Sistem Pendeteksi Intrusi”, 2012
- [9] Han J and Kimber M, *Data Mining Concepts and Techniques 2nd Edition*. San Francisco: Mofgan Kaufmann. 2006
- [10] F. Gorunescu, *Data Mining Concept, Models and Techniques*. Verlag Berlin: Heidelberg Springer. 2011
- [11] E. Prasetyo, *Data Mining Konsep dan Aplikasi Menggunakan Matlab.pdf*. Yogyakarta: Andi Offset. 2012