

Penerapan Data Mining Untuk Analisis Karakteristik DPT *Non-Participate* sebagai Prediksi Partisipan Pemilu dengan Menggunakan Metode *Naive Bayes Classifier*

Sofia Listyaningrum¹

^{1,3} Jurusan Teknik Informatika, FASILKOM UDINUS
Jln. Nakula 1 No 5-11 Semarang 50131 INDONESIA

¹111201105846@mhs.dinus.ac.id

Abstract— Election in Indonesia have been conducted since 1955 ago, but for the election of members of the executive body involving the public only, began in 2004. Participation of DPT (voters list) for the presidential election in 2014 decreased compared to 2009 (70.9%) is only 70.2%. This indicates that non-Participate DPT this year more than the previous election. To analyze the characteristics of the non-Participate DPT itself used data mining with classification techniques that if it has been known it can increase the level of community participation in the next election. The method used is Naive Bayes classifier, and the design of the study using the CRISP-DM. The research data is data DPT Presidential Election 2014 from Pasekaran village at Batang regency that will be evaluated using a confusion matrix and validated with split validation techniques. The highest accuracy results obtained after the data is converted research is 88.46%. In addition, the results also show the results of the analysis of non-Participate DPT characteristics of study data is the place of birth (Batang), age (17-25), marital status (unmarried), sex (male), occupation (student), address (outside RW 04 and RW 04).

Keywords— DPT non-participate, data mining, naive bayes classifier, data mining classification, CRISP-DM

I. PENDAHULUAN

Pemilihan umum (pemilu) bukan merupakan hal yang asing lagi bagi masyarakat Indonesia saat ini. Diantaranya adalah pilkada (pemilihan kepala daerah), pileg (pemilihan anggota legislatif), dan pilpres (pemilihan presiden). Negara Indonesia sendiri telah melaksanakan pemilihan umum pertama pasca kemerdekaan pada tahun 1955, dan pemilu 2014 ini merupakan yang ke-11[1]. Pemilihan anggota badan eksekutif baru dilaksanakan mulai tahun 2004, oleh karena itu masih banyak masalah yang terjadi dalam pemilihan tersebut terutama dalam hal DPT (Daftar Pemilih Tetap).

Dalam peraturan KPU No.26 Tahun 2013 dijelaskan bahwa DPT merupakan warga yang telah terdata untuk memilih pada tempat pemungutan suara (TPS) yang telah ditunjuk[1]. Menurut monitoring KPU dijelaskan ada beberapa masalah terkait DPT yaitu, KPU sulit mendapatkan NIK orang-orang yang berada di lembaga pemasyarakatan atau tahanan, pemilih pemula yang belum memiliki KTP yang sedang belajar baik di pesantren, asrama mahasiswa dan lain-lain di luar kota yang jumlahnya diperkirakan 3-5 persen dari NIK *invalid*, pemilih yang tidak memiliki identitas kependudukan, pemilih dengan KTP/KK lama dan NIK *invalid* sekitar 7-10 persen, dan pemilih yang sulit ditemui sekitar 5-8 persen dari NIK *invalid*.

Dengan adanya berbagai masalah tersebut, mengakibatkan kurangnya partisipasi masyarakat mengikuti pemilu ini.

Tingkat partisipasi pemilih dari tahun 1955-2009 menurut merdeka.com antara lain adalah, 1955 (91,4%), 1971 (96,6%), 1977 & 1982 (96,5%), 1987 (96,4%), 1992 (95,1%), 1997 (93,6%), 1999 (92,6%), 2004 (84,1%), 2009 (70,9%). Sedangkan berdasarkan hasil dari KPU persentase golput pilpres 2014 yaitu 29,8% yang berarti bahwa partisipasi masyarakat hanya 70,2%. Untuk desa Pasekaran Kab.Batang sendiri pada Pilpres tahun 2014 tingkat partisipasinya 73%, turun 1% jika dibandingkan dengan Pemilu sebelumnya pada tahun 2009 yang tingkat partisipasinya 74%.

Jika kemungkinan DPT yang akan tidak memilih (*non-participate*) diketahui lebih dini oleh KPU, maka akan lebih memudahkan di dalam menentukan target sosialisasi pemilu dan untuk menambah persentase partisipasi masyarakat daripada sebelumnya sebagai kontrol pemerintahan. Oleh karena itu dibutuhkan analisis untuk memprediksi DPT *non-participate* pada pemilu yang akan datang.

Data mining diartikan sebagai ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di *database* yang sangat besar[4]. Salah satu teknik *data mining* adalah teknik klasifikasi. Teknik klasifikasi adalah teknik pembelajaran untuk prediksi suatu nilai dari target variabel kategori[9]. Karena dalam prediksi DPT *non-participate* ini memiliki dua variable target yaitu *participate* dan *non-participate* maka penelitian ini akan menggunakan *data mining* teknik klasifikasi.

Naive bayes Classifier merupakan salah satu metode klasifikasi 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*.

Oleh karena itu dalam penelitian ini akan dilakukan analisis prediksi DPT *non-participate* dengan menggunakan metode klasifikasi, yaitu algoritma *naive bayes classifier*. Sedangkan untuk desain penelitian akan menggunakan model *CRISP-DM*.

II. STUDI PUSTAKA

2.1. Penelitian Terkait

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

1. Diana Tri Wahyuni, T.Sutojo, Ardytha Luthfiarta "Prediksi Hasil Pemilu Legislatif DKI Jakarta menggunakan *Naive Bayes* dengan Algoritma Genetika sebagai Fitur Seleksi".
2. Thimas Rusch, Ilro Lee, Kurt Hornik, Wolfgang Jank, Achim Zeiles "*Influencing Elections with Statistic Targeting Voters with Logistic Regression Trees*".
3. Gregg R.Murray, Chris Riley, Anthony Scime "*Pre-Election Polling: Identifying Likely Voters Using Iterative Expert Data Mining*".
4. Muhammad Fakhurrifqi, Retantyo Wardoyo "Analisis Perbandingan Algoritma *Nearest Neighbour*, C4.5 dan LVQ untuk Klasifikasi Kemampuan Mahasiswa".
5. Dwi Widiastuti "Analisis Perbandingan Algoritma SVM, *Naive Bayes*, dan *Decision Tree* dalam Mengklasifikasikan Serangan (*Attacks*) pada Sistem Pendeteksi Intrusi"

2.2. Tinjauan Pustaka

A. DPT (Daftar Pemilih Tetap)

Menurut peraturan KPU No 26 Tahun 2013 Bab II Pasal 6, telah diatur bahwa ada 4 pemilih pada Pemilu yaitu Daftar Pemilih Tetap (DPT), Daftar Pemilih Tetap Tambahan (DPTb), Daftar Pemilih Khusus (DPK), dan Daftar Pemilih Khusus Tambahan (DPKTb). DPT sendiri merupakan warga yang telah terdata untuk memilih pada tempat pemungutan suara (TPS) yang ditunjuk[7]. Untuk DPT *non-participate* sendiri merupakan DPT yang telah terdaftar tetapi tidak memilih pada saat diadakan pemilu. Banyaknya DPT *non-participate* mempengaruhi persentase partisipasi masyarakat dalam pemilu.

B. Data Mining

Dewasa 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[5]. 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 [9]. 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 merupakan salah satu metoda machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

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

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 [11]. 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

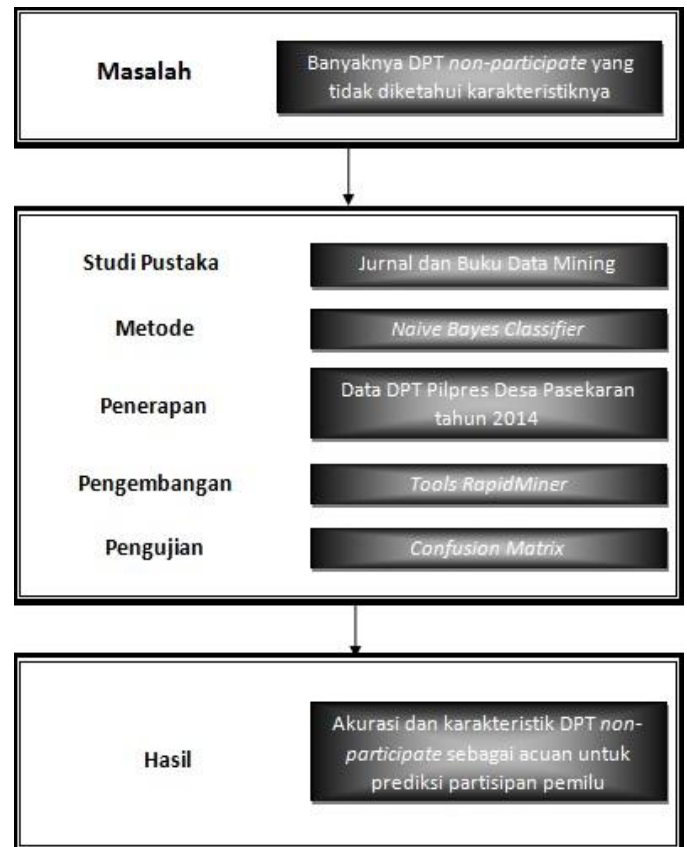
Classification	Predicted class	
	Class = Yes	Class = No
Class = Yes	a (<i>true positive-TP</i>)	b (<i>true negative-TN</i>)
Class = No	c (<i>false positive-FP</i>)	d (<i>false negative-FN</i>)

F. Split Validation

Split validation adalah teknik validasi yang membagi data menjadi dua bagian secara acak, sebagian menjadi *data training* sebagian lainnya menjadi *data testing*. Dengan menggunakan *split validation* akan dilakukan percobaan

training berdasarkan *split ratio* yang telah ditentukan sebelumnya, kemudian sisa dari *split ratio data training* dianggap sebagai *data testing*. *Data training* merupakan data yang akan digunakan dalam pembelajaran sedangkan *data testing* merupakan data yang belum pernah digunakan sebagai pembelajaran dan akan menjadi data pengujian kebenaran atau keakurasian hasil pembelajaran [12].

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 diperoleh dari *database* Daftar Pemilih Tetap (DPT) pemilihan presiden tahun 2014 yang didapatkan dari Komisi Pemilihan Umum (KPU) Kabupaten Batang.

3.1.2 Data Sekunder

Dalam penelitian juga digunakan data sekunder yang berfungsi untuk membantu penyusunan tugas akhir yang diantaranya adalah

e-book, jurnal, buku, beserta kumpulan materi yang membahas tentang *data mining*.

3.2 Metode Pengumpulan Data

3.2.1 Wawancara

Untuk wawancara dilakukan penulis dengan cara menanyakan langsung kepada petugas KPU tentang Daftar Pemilih Tetap (DPT) mengenai bagaimana partisipasi masyarakat dari tahun ke tahun, dan juga apa saja masalah yang timbul terkait DPT Kabupaten Batang saat ini.

3.2.2 Dokumentasi

Dokumentasi penelitian didapatkan setelah dilakukan wawancara sebelumnya. Dokumentasi berupa dokumentasi data yang antara lain seperti daftar DPT keseluruhan, diagram partisipasi dari pemilu-pemilu sebelumnya.

3.2.3 Studi Pustaka

Penulis juga menggunakan studi pustaka sebagai salah satu metode pengumpulan data yaitu dengan menggunakan buku, jurnal-jurnal, artikel, kumpulan materi yang berhubungan dengan DPT, data mining, dan *naive bayes classifier*.

3.3 Desain Penelitian

Dengan menggunakan CRISP-DM, penelitian ini memiliki beberapa langkah sebagai berikut :

1. Business Understanding

Penulis menggunakan data DPT Desa Pasekaran Kabupaten Batang, dimana DPT yang terdaftar berjumlah 3397 orang, tetapi yang memilih hanya 2487 orang. Tingkat partisipasinya adalah 73%, ada 910 orang yang tidak menggunakan hak pilihnya dalam pemilu.

2. Data Understanding

Data dalam penelitian ini menggunakan data yang diperoleh dari sumber data (data primer). Data tersebut didapat dari *database* DPT pemilu 2014 yang dimiliki oleh Komisi Pemilihan Umum (KPU) Kabupaten Batang. Data yang dikumpulkan adalah data DPT untuk desa Pasekaran Kabupaten Batang pada pemilu 2014 beberapa bulan yang lalu dengan atribut sebagai berikut :

Tabel 3.1 Atribut Data Penelitian

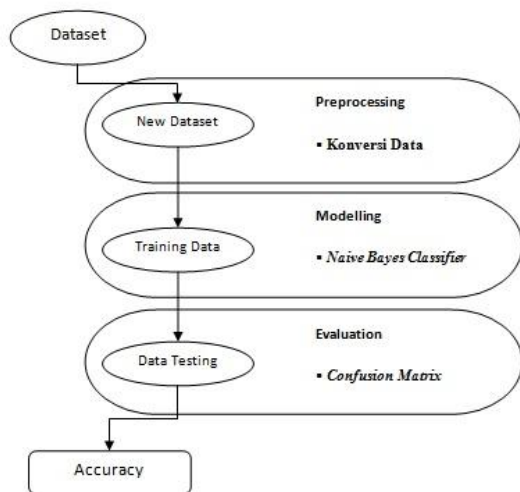
No. KK	atribut yang menginformasikan nomor kartu keluarga dari dpt
NIK	atribut yang menginformasikan nomor induk kependudukan dari dpt
Nama	atribut yang menginformasikan nama lengkap dpt
Tempat Lahir	atribut yang menginformasikan tempat lahir dari dpt
Tanggal Lahir	atribut yang menginformasikan tanggal lahir dari dpt
Umur	atribut ini menginformasikan umur dari dpt pada pemilu pilpres 2014
Status Perkawinan	atribut ini menginformasikan status perkawinan dari dpt, dimana S (Sudah Menikah), B (Belum Menikah), P (Pernah Menikah)
Jenis Kelamin	atribut ini menginformasikan jenis kelamin dpt, yaitu Lk untuk laki-laki dan Pr untuk perempuan
Alamat	atribut ini menginformasikan alamat dari dpt yang terdiri dari jalan/dukuh, RT tempat tinggal dpt dan juga RW dari alamat dpt tersebut
Pekerjaan	atribut ini menginformasikan pekerjaan dari dpt dari pelajar/mahasiswa, pegawai negeri sipil, buruh, pedagang, guru, karyawan swasta, karyawan BUMN, karyawan BUMD, ibu rumah tangga, pensiunan, wiraswasta, dokter, mekanik, belum/tidak bekerja, dll.
Keterangan	atribut ini menginformasikan keterangan tentang dpt yaitu <i>participate</i> apabila dpt memilih dan <i>non-participate</i> untuk dpt yang tidak memilih

3. Data Preparation

Dari 11 atribut pada *data understanding* hanya akan digunakan 7 atribut untuk data penelitian, yaitu tempat lahir, umur, status perkawinan, jenis kelamin, alamat, pekerjaan, dan keterangan sebagai label target.

4. Modelling

Metode yang akan digunakan dalam penelitian ini adalah Algoritma *Naive Bayes Classifier*. 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.

6. Penyebaran (*Deployment*)

Hasil dari penelitian ini berupa analisis yang mengarah ke *decision Support System* (DSS), yang diharapkan dapat digunakan oleh Komisi Pemilihan Umum (KPU) sebagai referensi untuk mengurangi DPT *non-participate* maupun untuk meningkatkan tingkat partisipasi pada pemilu yang akan datang dan juga bisa digunakan untuk rujukan pada penelitian selanjutnya.

IV HASIL PENELITIAN DAN PEMBAHASAN

4.1 Pengolahan Data

Data yang akan digunakan pada penelitian ini adalah data Daftar Pemilih Tetap (DPT) desa Pasekaran Kabupaten Batang pada pemilu 2014. Sumber data penelitian ini adalah dari Komisi Pemilihan Umum (KPU) Kabupaten Batang. Untuk data yang diperoleh dari KPU masih berupa *raw data* dimana data masih merupakan data murni dari KPU yang berisi atribut No.KK, NIK, nama, tempat lahir, tanggal lahir, umur, status perkawinan, jenis kelamin, alamat (Jalan/Dukuh,RT,RW), pekerjaan, keterangan *participate* atau *non-participate*. 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 tujuh (7) antara lain tempat lahir, umur, status perkawinan, jenis kelamin, alamat, pekerjaan, dan keterangan *participate* atau *non-participate*, sehingga ada

empat atribut yang tidak dipakai yaitu No.KK, NIK, nama, dan tanggal lahir. 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

Tempat Lahir	Umur	Status Perkawinan	Jenis Kelamin	Alamat	Pekerjaan	Keterangan
		B/S/P	Lk/Pr			
BATANG	17-25	B	Pr	RW 04	Pelajar/Mahasiswa	P
LUAR BATANG	17-25	B	Pr	RW 04	Pelajar/Mahasiswa	P
BATANG	17-25	B	Lk	RW 04	Pelajar/Mahasiswa	P
BATANG	17-25	B	Lk	RW 04	Pelajar/Mahasiswa	P
BATANG	17-25	B	Lk	RW 04	Pelajar/Mahasiswa	NP
BATANG	17-25	B	Lk	RW 04	Pelajar/Mahasiswa	P
BATANG	17-25	B	Lk	RW 04	Pelajar/Mahasiswa	NP
BATANG	17-25	B	Pr	RW 04	Pelajar/Mahasiswa	P
LUAR BATANG	17-25	B	Pr	RW 04	Pelajar/Mahasiswa	NP
BATANG	26-40	S	Lk	LUAR RW 04	Wiraswasta	P
LUAR BATANG	26-40	B	Lk	LUAR RW 04		NP
LUAR BATANG	41-55	S	Pr	LUAR RW 04	Belum/Tidak Bekerja	P
LUAR BATANG	41-55	B	Lk	LUAR RW 04	Wiraswasta	P

4.2 Pembahasan Hasil Percobaan

Percobaan pada penelitian ini menggunakan *tools Rapid Miner Ver 6.0.008*. Algoritma yang digunakan adalah *Naive Bayes Classifier* dan validasi yang dipakai adalah *split validation*, sedangkan 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 269 *record* dan 7 atribut menghasilkan tingkat akurasi tertinggi yaitu 88,46% yang sudah termasuk dalam rentang akurasi *good classification*. Selain itu karakteristik DPT *non-participate* yang didapat adalah tempat lahir (batang), umur (17-25), status perkawinan (belum menikah), jenis kelamin (laki-laki), pekerjaan (pelajar/mahasiswa), alamat (luar RW 04 dan RW 04).

Dari hasil karakteristik DPT *non-participate* yang sudah didapat, maka bisa dibuktikan dengan perhitungan manual menggunakan *naive bayes classifier* di bawah ini :

1. Menghitung jumlah kelas dari klasifikasi yang terbentuk

C1 = keterangan (*participate*)

C2 = keterangan (*non-participate*)

Perhitungan :

$$P(X|C_1) = P(\text{keterangan} = \text{"participate"}) = 214/269 = 0,796$$

$$P(X|C_2) = P(\text{keterangan} = \text{"non-participate"}) = 55/269 = 0,204$$

2. Menghitung jumlah kasus yang sama dari kelas X (tempat lahir=batang, umur=17-25, status=belum menikah, jenis kelamin=laki-laki, alamat=RW 04, pekerjaan=pelajar/mahasiswa)

Perhitungan :

$P(\text{tempat lahir} = \text{"batang"} \mid \text{keterangan} = \text{"participate"}) = 95/214 = 0,444$
 $P(\text{tempat lahir} = \text{"batang"} \mid \text{keterangan} = \text{"non-participate"}) = 34/55 = 0,618$

$P(\text{umur} = \text{"17-25"} \mid \text{keterangan} = \text{"participate"}) = 51/214 = 0,238$
 $P(\text{umur} = \text{"17-25"} \mid \text{keterangan} = \text{"non-participate"}) = 27/55 = 0,491$

$P(\text{status} = \text{"belum menikah"} \mid \text{keterangan} = \text{"participate"}) = 52/214 = 0,243$
 $P(\text{status} = \text{"belum menikah"} \mid \text{keterangan} = \text{"non-participate"}) = 39/55 = 0,709$

$P(\text{jenis kelamin} = \text{"laki-laki"} \mid \text{keterangan} = \text{"participate"}) = 98/214 = 0,458$
 $P(\text{jenis kelamin} = \text{"laki-laki"} \mid \text{keterangan} = \text{"non-participate"}) = 36/55 = 0,655$

$P(\text{alamat} = \text{"RW 04"} \mid \text{keterangan} = \text{"participate"}) = 211/214 = 0,986$
 $P(\text{alamat} = \text{"RW 04"} \mid \text{keterangan} = \text{"non-participate"}) = 54/55 = 0,982$

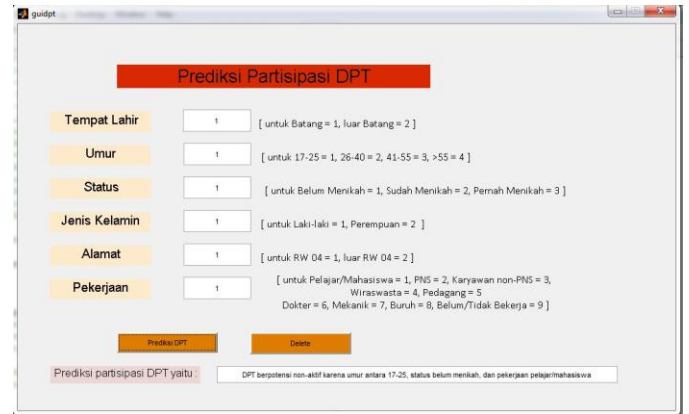
$P(\text{pekerjaan} = \text{"pelajar/mahasiswa"} \mid \text{keterangan} = \text{"participate"}) = 49/214 = 0,229$
 $P(\text{pekerjaan} = \text{"pelajar/mahasiswa"} \mid \text{keterangan} = \text{"non-participate"}) = 31/55 = 0,564$

- Mengkalikan semua hasil dari atribut
 $P(X \mid \text{keterangan} = \text{"participate"}) = 0,444 \times 0,238 \times 0,243 \times 0,458 \times 0,986 \times 0,229 = 0,0027$
 $P(X \mid \text{keterangan} = \text{"non-participate"}) = 0,618 \times 0,491 \times 0,709 \times 0,655 \times 0,982 \times 0,564 = 0,0780$
 $P(X \mid \text{keterangan} = \text{"participate"}) \times P(\text{keterangan} = \text{"participate"}) = 0,0027 \times 0,796 = 0,0021$
 $P(X \mid \text{keterangan} = \text{"non-participate"}) \times P(\text{keterangan} = \text{"non-participate"}) = 0,0780 \times 0,204 = 0,0159$

- Bandungkan hasil kelas :
 $P(X \mid \text{keterangan} = \text{"participate"}) \times P(\text{keterangan} = \text{"participate"})$
 $P(X \mid \text{keterangan} = \text{"non-participate"}) \times P(\text{keterangan} = \text{"non-participate"})$

Kesimpulan akhir : dengan karakteristik DPT tersebut maka DPT digolongkan dalam klasifikasi keterangan="non-participate" karena hasil probabilitas non-participate lebih besar dari participate.

4.3 Implementasi pada Matlab



Gambar 4. 1 Tampilan pada GUI Matlab

Pada tampilan GUI (*Graphical User Interface*) gambar 4.12 diatas, aplikasi dijalankan dengan memasukkan inputan berupa nomor sesuai dengan keterangan yang ada. Inputannya antara lain adalah tempat lahir, umur, status, jenis kelamin, alamat, dan pekerjaan. *Button* Prediksi DPT digunakan untuk melihat prediksi partisipasi DPT yaitu berpotensi non-aktif atau aktif sesuai inputan. Sedangkan *button delete* digunakan untuk menghapus inputan dan hasil prediksi, apabila ingin melakukan inputan kembali.

V KESIMPULAN DAN SARAN

a. Kesimpulan

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

- Dari atribut yang digunakan pada penelitian, diketahui karakteristik DPT *Non-Participate* dari data yang didapat peneliti adalah tempat lahir (batang), umur 17-25, status perkawinan B (belum menikah), jenis kelamin Lk (laki-laki), pekerjaan pelajar/mahasiswa, sedangkan untuk alamat antara Luar RW 04 dan RW 04 seimbang dengan DPT *participate*.
- Cross-Industry Standart Process for Data Mining (CRISP-DM)* dengan algoritma *Naive Bayes Classifier (NBC)* dapat diterapkan untuk mengolah data DPT (Daftar Pemilih Tetap) yang digunakan untuk mengetahui karakteristik DPT *non-participate*.
- Tingkat akurasi data setelah dikonversi yang didapatkan setelah dilakukan evaluasi dengan *confusion matrix* adalah 88,46% yang sudah merupakan rentang akurasi *good classification*.

b. Saran

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

1. Untuk mengembangkan penelitian bisa dilakukan penambahan *record* data untuk mendapatkan nilai akurasi yang lebih baik.
2. Dapat dilakukan pengembangan dengan metode klasifikasi selain algoritma *naive bayes classifier*.
3. Dengan diketahuinya karakteristik DPT *non-participate* diharapkan bisa menjadi acuan KPU sebagai sasaran sosialisasi agar partisipasi di pemilu yang akan datang lebih tinggi dari pemilu sebelumnya.
4. Melakukan pengembangan penelitian dengan DPT dari daerah yang berbeda sehingga lebih bisa mengetahui karakteristiknya secara global.

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] L Agustino and M.A Yusoff, "Jurnal Kajian Politik dan Masalah Pembangunan," *Jurnal Enam Bulanan*, vol. 5, no. 1, p. 416, 2009.
- [2] Asdhar. (2014, February) Luwuraya.com. [Online]. HYPERLINK "http://www.luwuraya.com" <http://www.luwuraya.com>
- [3] G.R Murray, C Riley, and A Scime, "Pre-Election Polling: Identifying Likely Voters Using Iterative Expert Data Mining," 2009.
- [4] Yudho Giri Sucahyo. (2003) IlmuKomputer.Com.
- [5] 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.
- [6] Tia Subekti, "Partisipasi Politik Masyarakat dalam Pemilihan Umum," *Studi Turn of Voter dalam Pemilihan Umum Kepala Daerah Kabupaten Magetan Tahun 2013*, 2014.
- [7] Larose D T, *Discovering Knowledge in Database*. New Jersey: John Wiley and Sons Inc, 2005.
- [8] M.J.A Berry and G.S Linoff, *Data Mining Techniques for Marketing, Sales, Customer Relationship Management*, 2nd ed. United States of America: Wiley Publishing Inc, 2004.
- [9] Kusrini and E.T Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- [10] A Basuki and I Syarif, "Decision Tree," Politeknik Elektronika Negeri Surabaya - ITS, Surabaya, 2003.
- [11] F. Gorunescu, *Data Mining Concept, Models and Techniques*. Verlag Berlin: Heidelberg Springer, 2011.
- [12] I.H Witten, E Frank, and M.A Hall, *Data Mining Practical Machine Learning Tools and Technique*. Burlington: Morgan Kaufmann Publisher, 2011.
- [13] Diana Tri Wahyuni, T Sutojo, and Ardytha Luthfiarta, "Prediksi Hasil Pemilu Legislatif DKI Jakarta Menggunakan Naive Bayes dengan Algoritma Genetika sebagai Fitur Seleksi," 2014.
- [14] Thimas Rusch, Ilro Lee, Kurt Homik, Wolfgang Jank, and Achim Zeleis, "Influencing Elections with Statistic Targeting Voters with Logistic Regression Trees," 2013.
- [15] Muhammad Fakhurrizzqi and Retantyo Wardoyo, "Analisis Perbandingan Algoritma Nearest Neighbour, C4.5, dan LVQ untuk Klasifikasi Kemampuan Mahasiswa," 2013.
- [16] Dwi Widiastuti, "Analisis Perbandingan Algoritma SVM, Naive Bayes, dan Decision Tree dalam Mengklasifikasikan Serangan (Attacks) pada Sistem Pendeteksi Intrusi," 2012.