

# Algoritma Naive Bayes untuk Prediksi Kepuasan Pelayanan Perekaman e-KTP

## (Naive Bayes Algorithm for Satisfaction Prediction of e-ID Card Recording Service)

Tri Herdiawan Apandi<sup>1</sup>, Castaka Agus Sugianto<sup>2</sup>

<sup>1</sup>Program Studi Manajemen Informasi, Politeknik Negeri Subang, Subang

<sup>2</sup>Program Studi Teknik Informatika, Politeknik TEDC Bandung, Cimahi, 40513

<sup>1</sup>h.apandi@gmail.com

<sup>2</sup>castaka@poltektedc.ac.id

**Abstrak**— Layanan publik adalah semua kegiatan layanan yang dilakukan oleh penyedia layanan publik sebagai upaya untuk memenuhi kebutuhan penerima layanan sesuai dengan ketentuan dan aturan yang dibuat. Pelayanan e-KTP sebagai salah satu layanan publik yang masih belum maksimal, hal ini dibuktikan dengan data yang telah dikumpulkan. Sampel data yang diambil adalah data dari Kecamatan Batujajar Kabupaten Bandung Barat. Kecamatan Batujajar menjadi tempat pengambilan data karena masih banyak data yang belum terekam sebanyak sebanyak 7974 data e-KTP. Untuk mengetahui lebih dini tingkat kepuasan masyarakat terhadap pelayanan perekaman e-KTP, salah satunya menggunakan data mining. Pada penelitian ini akan memprediksi tingkat kepuasan pelayanan perekaman e-KTP dengan mengumpulkan 17 indikator dengan menggunakan skala *likert*. Data kepuasan yang telah dikumpulkan akan dipisahkan antara data latih dan data uji. Hasil model data latih akan digunakan untuk melihat akurasi pada data uji. Dari hasil pengujian, ditemukan bahwa algoritma naive bayes memiliki akurasi sebesar 91.70%. Adapun algoritma pembandingan dalam hal ini Decision Tree, yaitu sebesar 65.90%.

**Kata-kata kunci:** *Naive Bayes, Decision Tree, e-KTP, Klasifikasi.*

**Abstract**— Public services are all service activities carried out by community service providers to meet the needs of service recipients, as well as in the implementation of rules. The e-ID Card (e-KTP) service as one of the public services provided is still not maximal, it is proven by the data that has been collected. Samples of data taken are data from Batujajar District, West Bandung Regency. Batujajar District is the place for data collection because there are 7974 e-ID Card data that have not been recorded. To find out the level of community satisfaction with e-ID Card recording services, one of them uses data mining. The study will predict the level of satisfaction of e-ID Card recording

services by collecting 17 indicators using a Likert scale. Satisfaction data that has been gathered will be collected between training data and test data. The results of the training data model will be used to see the accuracy of the test data. From the results of testing, the Bayes algorithm has an accuracy of 91.70%. There is a comparison algorithm in this case, the Decision Tree, which is 65.90%.

**Keywords:** *Naive Bayes, Decision Tree, e-ID Card, Classification.*

### I. PENDAHULUAN

Layanan publik dilakukan oleh penyedia layanan publik untuk memenuhi kebutuhan penerima layanan dan untuk mematuhi persyaratan hukum Menteri Pemberdayaan Aparatur Negara 2004 [1]. Layanan Publik tidak dapat dimaksimalkan kecuali dengan cara bekerja secara optimal terutama pejabat publiknya. Menurut Surat edaran Menteri Dalam Negeri No. 471/1768 / SJ berdasarkan catatan yang tersisa di Kabupaten Bandung Barat adalah 7.974 orang, yang belum terekam harus diselesaikan paling lambat 30 September 2016. Sedangkan e-KTP yang belum tercetak menunjukkan angka di 27.507 orang yang harus selesai pada 31 desember. Kartu e-KTP wajib 2016 berjumlah 1.014.742 jiwa [2].

Salah satu dari kecamatan yang memiliki masalah pencatatan data e-KTP di Kabupaten Bandung Barat adalah Kecamatan Batujajar. Menurut hasil observasi sementara masalah yang didapat adalah waktu pelayanan yang relative lama, kurangnya sosialisasi tentang pembuatan e-KTP, kurangnya fasilitas dan kelengkapan pembuatan e-KTP seperti komputer *finger print*, terjadi antrian yang panjang ketika pengurusan e-KTP. Akibatnya masih banyak warga yang belum mengurus kartu identitas tersebut. Masalah yang sudah dijelaskan di atas membutuhkan proses penanganan

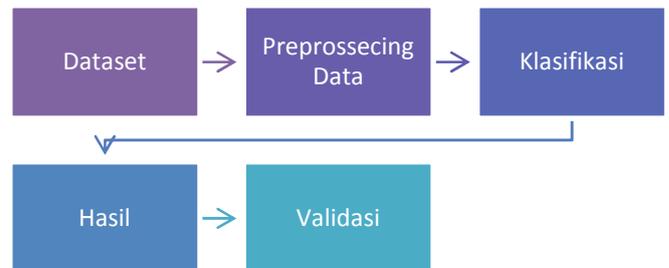
lebih awal untuk melihat tingkat kepuasan pelayanan pembuatan e-KTP. Cara untuk melihat dengan cepat tingkat kepuasan pelayanan perekaman E-KTP adalah dengan melakukan prediksi pada data kepuasan pelayanan perekaman E-KTP. Salah satu metode untuk melakukan prediksi tersebut dengan menggunakan *data mining*.

Beberapa penelitian data mining terutama klasifikasi sering digunakan dalam memprediksi tingkat kepuasan pelayanan adapun beberapa peneliti terkait tentang pelayanan diantaranya: penelitian yang dilakukan oleh Fahmi Rezha, Siti Rochmah dan Siswidiyanto mengenai pelayanan perekaman kartu tanda penduduk elektronik di Kota Depok hasilnya 75,8% variabel Kepuasan masyarakat akan dipengaruhi oleh kualitas pelayanan[3]. Sedangkan yang dilakukan Castaka Agus S dan Tri Herdiawan Apandi mengenai kepuasan pelayanan e-KTP menggunakan algoritma genetika dengan hasil akurasi 85.64%[4]. Berikutnya yang dilakukan oleh firdaus hamta dalam mengukur tingkat kepuasan pelayanan di samsat batam menggunakan algoritma naïve bayes dengan hasil akurasi 94.04%[5]. Dalam penelitian ini menggunakan algoritma Naïve Bayes, karena berdasarkan hasil penelitian terdahulu algoritma Naive Bayes adalah salah satu algoritma terbaik ketika dibandingkan dengan algoritma lain, seperti algoritma *logistic regression*, *neural network*, *support vector machine* dan *k-nearest neighbour*. Membuktikan bahwa *Naïve Bayes* menjadi Algoritma yang terbaik dibandingkan dengan Algoritma *Decesion Tree*, selain itu akan dilakukan *T-test* untuk melihat signitikan tidaknya dari kedua algoritma yang diuji. *Naïve Bayes* adalah algoritma klasifikasi yang cukup sederhana dan mudah diimplementasikan [6] sehingga algoritma ini sangat efektif ketika diuji dengan dataset yang benar, terutama jika *Naïve Bayes* dikombinasikan dengan pemilihan fungsi, sehingga *Naïve Bayes* dapat mengurangi redundan pada data [7], selain itu *Naïve Bayes* menunjukkan hasil yang bagus ketika digabungkan dengan metode clustering [8]. *Naïve Bayes* terbukti memiliki akurasi yang tinggi dibandingkan dengan support vector machine[9]. Hal ini sejalan dengan yang dilakukan untuk memprediksi data customer dengan metode naïve bayes memiliki hasil yang baik [10]. *Naïve Bayes* adalah algoritma klasifikasi dalam teknik data mining yang didalamnya menggunakan teori bayes [11]. Pada proses klasifikasi/prediksi *Naïve Bayes* sangat baik, pada penelitian sebelumnya dilakukan eksperimen terhadap 20 kelompok berita yang masing-masing memiliki data yang seimbang didapat naïve bayes mendapatkan penambahan akurasi yang baik

dibandingkan dengan SVM[12]. Naïve Bayes ini sangat familiar sehingga banyak digunakan, selain itu mempunyai performa yang bagus pada banyak domain. Pada penelitian terdahulu memakai domain rating film untuk diprediksi. Hasil rata-rata akurasi prediksi rating film adalah rendah. [13].

## II. METODE

Pada penelitian ini metode dapat dilihat pada Gambar 1.



Gambar 1. Metode yang diusulkan

Berikut ini adalah penjelasan dari Gambar 1. yang dilakukan untuk mendapat prediksi kepuasan pelayanan E-KTP:

### A. Dataset

*Dataset* yang dikumpulkan untuk prediksi kepuasan layanan e-KTP diambil dari data kuisioner yang telah disebar. Kuisioner diambil dari Kecamatan Batujajar, data yang berhasil dikumpulkan sebanyak 3977 data. Data tersebut dikumpulkan dan kemudian akan diolah. Kecamatan Batujajar menjadi tempat pengambilan data karena masih banyak data yang belum terekam sebanyak sebanyak 7974 data e-KTP.

Indikator yang digunakan dalam pembuatan kuisioner pada penelitian menggunakan indikator yang digunakan oleh Direktorat Jenderal Sumber Daya dan Perangkat Pos dan Informatika (Ditjen SDPPI). Adapun indikatornya sebagai berikut: Prosedur pelayanan, Persyaratan pelayanan, Kejelasan petugas pelayanan, Kesopanan dan keramahan petugas, kenyamanan dan keamanan lingkungan, kemampuan petugas pelayanan, kedisiplinan petugas pelayanan, tanggung jawab petugas pelayanan, kenyamanan masyarakat dalam berinteraksi dengan petugas, kecepatan dan ketepatan pelayanan, respon terhadap keluhan dan saran masyarakat, kepastian jadwal pelayanan, kepastian biaya pelayanan, kewajaran biaya pelayanan, keadilan mendapatkan pelayanan, kepedulian unit layanan terhadap masyarakat, Kesungguhan petugas unit layanan dalam membantu masyarakat. Adapun skala

pengukuran yang dipakai dalam kuisioner ini menggunakan Skala Likert.

**B. Preprocessing Data**

Proses preparasi yang dilakukan adalah dengan mengubah dari data kuisioner kedalam format excel, kemudian dilakukan analisis, identifikasi, memperbaiki data yang tidak konsisten dan membersihkan dataset dari noisy, duplikasi data. Setelah melakukan proses ini dataset siap diolah kedalam *Machine Learning*.

**C. Klasifikasi**

Klasifikasi yang digunakan adalah *Naïve Bayes* dan *Decision Tree* untuk membandingkan kedua algoritma tersebut. Klasifikasi dari dataset ini adalah proses untuk menentukan pelayanan tersebut.

**D. Hasil Output**

*Output* yang dihasilkan oleh *Naïve Bayes* dan *Decision Tree* adalah akurasi dari pengujian yang dilakukan.

**E. Validasi**

Validasi dilakukan menggunakan *Cross Validation*. *Cross-validation* metode digunakan untuk memprediksi keakuratan data pengujian sedangkan untuk melihat perbedaan antara *Naïve Bayes* dan *Decision Tree*. Selain menggunakan *Cross Validation* pada penelitian ini menggunakan T-Test dalam penelitian ini direncanakan nilai harus dibawah 0.050.

III. HASIL DAN PEMBAHASAN

**A. Hasil Pengujian Algoritma Decision Tree**

Berikut adalah hasil pengujian menggunakan algoritma *decision tree*. dari tabel I menunjukkan bahwa algoritma *decision tree* mendapatkan hasil akurasi sebesar 65,90%, *Precision* 73,44%, *Recall* 2,11%, AUC 0,602 dan *Execution Time* 57 second (Tabel 1).

TABEL I  
HASIL ALGORITMA DECISION TREE

|                     | <i>true</i> Puas | <i>true</i> Cukup Puas | <i>Class Precision</i> |
|---------------------|------------------|------------------------|------------------------|
| pred. Puas          | 2592             | 1346                   | 65,82%                 |
| pred. Cukup Puas    | 10               | 29                     | 74,36%                 |
| <i>Class Recall</i> | 99,62%           | 2,11%                  |                        |

**B. Hasil Pengujian Algoritma Naïve Bayes**

Berikut adalah hasil pengujian menggunakan algoritma *Naïve Bayes*. dari tabel II menunjukkan bahwa algoritma *Naïve Bayes* mendapatkan hasil akurasi

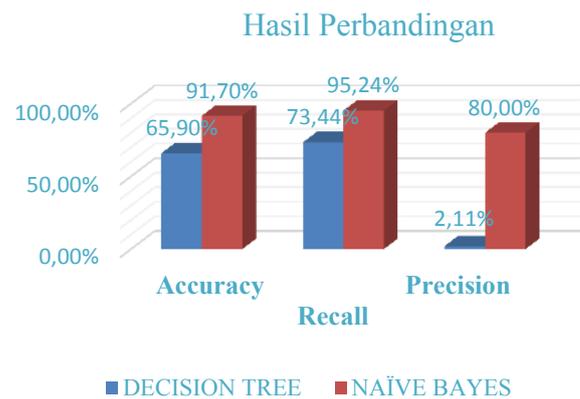
sebesar 91,70%, *Precision* 95,24%, *Recall* 80,00%, AUC 0,982 dan *Execution Time* 0 second seperti tersaji pada Tabel 2.

TABEL II  
HASIL ALGORITMA NAÏVE BAYES

|                     | <i>true</i> Puas | <i>true</i> Cukup Puas | <i>Class Precision</i> |
|---------------------|------------------|------------------------|------------------------|
| pred. Puas          | 2547             | 275                    | 90,26%                 |
| pred. Cukup Puas    | 55               | 1100                   | 95,24%                 |
| <i>Class Recall</i> | 97,89%           | 80,00%                 |                        |

**C. Hasil Perbandingan Algoritma**

Hasil perbandingan dari algoritma *decision tree* dan algoritma *Naïve Bayes* bisa dilihat pada Gambar 2 pada gambar tersebut. Algoritma *Decision Tree* menghasilkan pohon keputusan dimana pohon tersebut berguna untuk membuat aturan pada data yang baru, sedangkan pada algoritma *Naïve Bayes* akan membuat aturan yang berdasarkan kemungkinan yang bersyarat atau dapat digunakan pada data yang bias. Dari hasil gambar 2 sangat jelas terlihat bahwa algoritma *naïve bayes* lebih baik dibandingkan dengan algoritma *decision tree*. algoritma *decision tree* mempunyai tingkat akurasi sebesar 65,90% dan algoritma *naïve bayes* dan menggunakan *cross validation* 91,70%, *decision tree precision* 73,44% sedangkan *naïve bayes precision* 95,24%, *decision tree recall* 2,11% sedangkan *naïve bayes recall* 80,00%, *decision tree AUC* 0.602 sedangkan *AUC naïve bayes* 0.982. dan dalam segi waktu eksekusi algoritma *decision tree* menghabiskan waktu 57 second sedangkan *naïve bayes* menghabiskan waktu eksekusi 0 second.



Gambar 2. Hasil perbandingan accuracy, precision, dan recall

**D. Hasil Pengujian T-Test**

Pengujian menggunakan *t-test* ditujukan untuk mengetahui apakah data yang telah diolah oleh *rapid miner* menggunakan algoritma *Decision Tree* dan algoritma pembandingan yaitu *Naïve Bayes* dan menggunakan *Cross Validation* hasilnya signifikan atau tidak, dalam teori uji t dikatakan signifikan jika nilai  $\alpha=0.050$ . Dalam hasil penelitian ini didapat nilai  $\alpha=0.000$ , jadi dalam penelitian ini hasilnya signifikan. Dikatakan signifikan bisa dilihat pada Gambar 3.

|                 |                 |                 |
|-----------------|-----------------|-----------------|
|                 | 0.657 +/- 0.006 | 0.917 +/- 0.020 |
| 0.657 +/- 0.006 |                 | <b>0.000</b>    |
| 0.917 +/- 0.020 |                 |                 |

Gambar 3. Hasil T-Test

#### IV. PENUTUP

Berdasarkan hasil pengujian yang dilakukan maka dapat menarik kesimpulan dari keseluruhan pengujian yang dilakukan *Algoritma Naïve Bayes* lebih akurat untuk data tingkat kepuasan pelayanan e-ktip di Kecamatan Batujajar dengan tingkat akurasi sebesar 91.70% dan nilai pengujian *f-measure* sebesar 93,92%, sedangkan hasil dari akurasi algoritma *decision tree* adalah 65.90% dan nilai pengujian *f-measure* adalah 79,26% dan *execution time* dari algoritma *naïve bayes* 0 *second* dan dari algoritma *decision tree* 57 *second*.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Ditjen Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi, dan Pendidikan Tinggi, sebagai pemberi hibah kompetitif nasional yang telah memberi dukungan dana terhadap penelitian ini.

#### DAFTAR PUSTAKA

[1] Menteri Pendayagunaan Aparatur Negara, "Keputusan Menteri Pendayagunaan Aparatur Negara Nomor :

Kep/25/M.Pan/2/2004 Tentang Pedoman Umum Penyusunan Indeks Kepuasan Masyarakat Unit Pelayanan Instansi Pemerintah," p. 23, 2004.

- [2] Pojokjabar.com, "Disdukcapil Kabupaten Bandung Barat Kebut Pencetakan E-KTP," 2016. .
- [3] F. Rezha, S. Rochmah, and Siswidiyanto, "Analisis Pengaruh Kualitas Pelayanan Publik terhadap Kepuasan Masyarakat(Study Tentang Pelayanan Perekaman Kartu Tanda Penduduk Elektronik (e-KTP) di Kota Depok)," p. 10, 2016.
- [4] C. A. Sugianto and T. H. Apandi, "Algoritma Genetika untuk Optimalisasi Klasifikasi Kepuasan Pelayanan e-KTP," *J. Sink. J. Penelit. Tek. Inform.*, vol. 3, no. 1, pp. 186–191, 2018.
- [5] F. Hamta, "Analisis Penerapan Data Mining Dalam Mengukur Tingkat Kepuasan Masyarakat Pada Pelayanan Samsat Batam," pp. 1–17, 2017.
- [6] Fitriyani and R. S. Wahono, "Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naïve Bayes," *J. Softw. Eng.*, vol. 1, no. 2, pp. 101–108, 2015.
- [7] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools and Techniques*. Elsevier, 2011.
- [8] C. A. Sugianto, "Analisis Komparasi Algoritma Klasifikasi Untuk Menangani Data Tidak Seimbang Pada Data Kebakaran Hutan," *Techno.com*, vol. 14, no. 4, pp. 336–342, 2015.
- [9] T. H. Apandi and C. A. Sugianto, "Analisis Komparasi Machine Learning Pada Data Spam Sms," *J. TEDC*, vol. 12, no. 1, pp. 58–62, 2018.
- [10] S. De Cnudde and D. Martens, "Loyal to your city? A data mining analysis of a public service loyalty program," *Decis. Support Syst.*, vol. 73, pp. 74–84, 2015.
- [11] B. Santoso, *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu, 2007.
- [12] S. Hassan and M. Rafi, "Comparing SVM and Naïve Bayes Classifiers for Text Categorization with Wikitology as knowledge enrichment," in *Multitopic Confernce (INMIC), IEEE 14th Internasional*, 2012.
- [13] R. W. Pratiwi and Y. S. Nugroho, "Prediksi Rating Film Menggunakan Metode Naïve Bayes," *J. Tek. Elektro*, vol. 8, no. 2, pp. 60–63, 2017.