

Klasifikasi Status Kesejahteraan Rumah Tangga di Kabupaten Siak Menggunakan Algoritma Naive Bayes Classifier

Arianto Tarigan¹, Mustakim², Ervan Wahyudi³, Jeni Adhiva⁴

^{1,2,3,4} Puzzle Research Data Technology Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau,

^{1,2,3,4} Program Studi Sistem Informasi Fakultas Sains dan Teknologi
Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl.HR. Soebrantas No. 155 Simpang Baru, Tampan, Pekanbaru, Riau – Indonesia 28293

e-mail: arianto.tarigan@students.uin-suska.ac.id¹ mustakim.uin-suska@gmail.com²

ervan.wahyudi24@students.uin-suska.ac.id³ jeniadhiva@gmail.com⁴

Abstrak

Dinas Sosial Kabupaten Siak dalam menentukan status kesejahteraan keluarga berdasarkan musyawarah dengan menentukan kesejahteraan berdasarkan aspek kesehatan dan pendidikan. Oleh karena itu, perlu adanya sebuah sistem klasifikasi yang mendukung pihak Dinas Sosial dalam menentukan status kesejahteraan keluarga di Kabupaten Siak secara efektif dan tepat. Penelitian ini dimaksudkan untuk melakukan klasifikasi status kesejahteraan rumah tangga menggunakan algoritma Naive Bayes Classifier. Atribut yang digunakan pada penelitian ini adalah NIK, kecamatan, jenis cacat, penyakit kronis, ijazah tertinggi, lapangan usaha dari pekerjaan utama, status kedudukan dalam Pekerjaan Utama, Status bangunan tempat tinggal, status lahan tempat tinggal, jenis lantai, jenis dinding, jenis atap, sumber air minum, sumber penerangan utama, dan jumlah anggota keluarga. Penentuan atribut ini telah dikonfirmasi kepada pihak Dinas Sosial Kabupaten Siak melalui wawancara sebagai fitur utama penentu status kesejahteraan rumah tangga. Penelitian melakukan pembagian data latih dan data uji dengan menggunakan k-fold cross validation dilakukan dengan 4 cross dan 10 kali percobaan menggunakan tools Weka dan data kesejahteraan tahun 2017 sebanyak 19.612 data. Hasilnya didapatkan akurasi tertinggi pada k4 dengan akurasi rata-rata 53%. Hasil dari perhitungan algoritma tersebut akan diimplementasikan kedalam sebuah sistem informasi yang diharapkan mampu memberikan pemecahan masalah sehingga dapat membantu pihak Dinas Sosial dalam menentukan status kesejahteraan secara efektif dan tepat.

Kata kunci: cross validation, kesejahteraan keluarga, klasifikasi, naive bayes classifier.

Abstract

Social Office of Siak Regency for making certain the classifications of welfare household status based on a discussion by resolving the welfare based on the education and health aspect. Hence, we need classification to support Social Office setting on the family welfare status in Siak District correctly and effectively. This research purpose is setting on the classifications of the household welfare status by using Naive Bayes Classifier (NBC) Algorithm. Instruments in this research are using Identity card, regency, type of disability, kind of chronic ill, the last education, the main job, the position of the main job, status of the house building, status of the house land, type of the floor, type of the wall, type of the roof, source of the drinking water, source of the house light and number of family member. These instruments have been confirmed by Social office in Siak District through an interview as the main determinant of the family welfare status. There are two data in this research called training data and testing data by using k-fold cross-validation, we've done through 4 cross and 10 times trial by using weka tools and the welfare data in 2017 as much as 19.612 data. The result of highest algorithm calculation on K4 has an average accuracy of around 53%. That result is going to implement into the information system which hoped be able to give a problem resolving so that helping Social Office to determine the welfare status effectively and properly.

Keywords: cross validation, family welfare, classification, naive bayes classifier.

1. Pendahuluan

Kesejahteraan adalah suatu adat kehidupan dan penghidupan sosial, material, maupun spiritual yang meliputi rasa keselamatan, kesusilaan dan ketentraman lahir batin yang memungkinkan setiap masyarakat untuk melakukan usaha dalam memenuhi kebutuhan jasmani, rohani dan sosial yang sebaik-baiknya bagi diri, rumah tangga serta masyarakat [1]. Kesejahteraan sosial dalam arti luas mencakup berbagai tindakan yang dilakukan manusia untuk mencapai tingkat kehidupan masyarakat yang lebih baik. Kemiskinan sendiri merupakan salah satu masalah pembangunan kesejahteraan sosial [2]. Bagian penting untuk mendukung strategi penanggulangan kemiskinan adalah tersedianya data kemiskinan yang akurat dan tepat sasaran [3].

Kabupaten Siak merupakan kabupaten terkaya kelima di Indonesia dengan kekayaan 993.20 miliar berdasarkan laporan Badan Pemeriksa Keuangan tahun 2015 melalui situs website radar Pekanbaru. Dengan kekayaan alam yang berlimpah berdasarkan data dari Dinas Sosial Kabupaten Siak masyarakat Siak masih banyak yang tergolong ke dalam keluarga belum sejahtera. Berdasarkan data Badan Pusat Statistik (BPS) Kota Pekanbaru dalam website-nya menyebutkan bahwa jumlah penduduk di Kabupaten Siak pada tahun 2016 mencapai 453.052 jiwa dengan jumlah penduduk kategori sejahtera Berdasarkan data Tim Nasional Program Penanganan Kemiskinan (TNP2K) dalam website-nya juga menyebutkan bahwa jumlah penduduk sejahtera di Kabupaten Siak pada tahun 2016 mencapai 87.059 jiwa dengan 22.050 jiwa Kartu Keluarga. Jumlah penduduk tersebut juga Berdasarkan Kepmensos Nomor 57/HUK/2016.

Secara Keilmuan, Dinas Sosial Kabupaten Siak kurang tepat dalam menentukan status kesejahteraan keluarga, dilihat dari aturan penentuan kesejahteraan menurut Cahyat terdiri dari aspek kesehatan, aspek pendidikan dan aspek ekonomi. Namun pada kenyataannya aspek ekonomi tidak diperhitungkan oleh pemerintah setempat. Mereka menentukan tingkat kesejahteraan berdasarkan musyawarah dengan menentukan kesejahteraan berdasarkan aspek kesehatan dan pendidikan. Oleh karena itu, perlu adanya sebuah sistem klasifikasi yang mendukung pihak dinas sosial dalam menentukan status kesejahteraan keluarga di Kabupaten Siak secara efektif dan tepat.

Ketepatan penentuan status kesejahteraan rumah tangga dapat diklasifikasikan dengan menggunakan teknik data mining. Secara teknis, data mining merupakan sebuah proses untuk menemukan hubungan antara banyak bidang dalam sebuah dataset besar [4]. Salah satu metode yang ada dalam data mining ialah klasifikasi. Yang mana metode tersebut akan digunakan dalam penelitian ini. Salah satu algoritma data mining dengan metode klasifikasi adalah Naive Bayes Classifier (NBC).

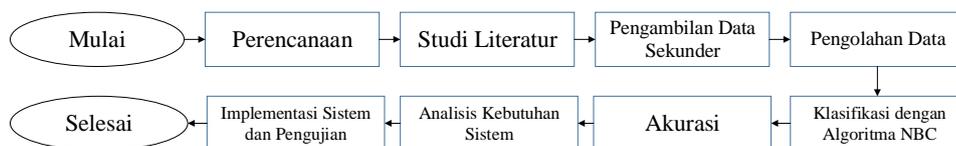
Penelitian terdahulu yang dilakukan tentang klasifikasi rumah layak huni yang dilakukan oleh Simatupang pada tahun 2016, penelitian ini mempunyai akurasi 71,43% dan tingkat error sebesar 28,57% dengan menggunakan algoritma Learning Vector Quantization. Sedangkan penelitian menggunakan algoritma Naive Bayes Classifier mempunyai tingkat akurasi 95,24% dan tingkat error 4,76% [5]. Kemudian penelitian selanjutnya tentang Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes Classifier yang dilakukan oleh Nugraha dkk pada tahun 2017 dengan data pasien yang diambil antara tahun 2014-2015 pada Rumah Sakit Umum Daerah Undata Palu memiliki akurasi yang sangat tinggi yaitu 90% [6]. Penelitian berikutnya tentang algoritma Naive Bayes Classifier yang dilakukan oleh Karyadiputra pada tahun 2016, penelitian ini membahas tentang Klasifikasi Status Kesejahteraan Rumah Tangga Keluarga Binaan Sosial menggunakan algoritma Naive Bayes menunjukkan bahwa tingkat akurasi yang cukup baik yaitu sebesar 85,80 % [7]. Kemudian penelitian berikutnya tentang Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining menggunakan algoritma ID3 dan Naive Bayes menunjukkan akurasi yang cukup rendah yaitu sebesar 62% untuk ID3 dan 65% untuk Naive Bayes [8].

Berdasarkan uraian di atas dan didukung oleh beberapa penelitian sebelumnya serta acuan dasar penetapan kesejahteraan yang dikemukakan oleh Cahyat, maka pada penelitian ini akan membangun sebuah sistem informasi untuk mengklasifikasi status kesejahteraan rumah tangga yang dapat digunakan sebagai pengetahuan yang bermanfaat dan dapat membantu dalam menentukan status kesejahteraan secara cepat dan akurat oleh Dinas Sosial Kabupaten Siak.

2. Metodologi Penelitian

Data mining adalah sebuah istilah yang digunakan untuk menggali sebuah pengetahuan baru di dalam database besar. Data mining juga berguna untuk mengidentifikasi informasi yang bermanfaat dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning dalam sebuah dataset besar [9].

Adapun diagram alur pada penelitian ini sebagai berikut:



Gambar 1. Metodologi Penelitian

2.1. Klasifikasi

Klasifikasi adalah penempatan objek-objek ke salah satu dari beberapa kategori yang telah ditetapkan sebelumnya. Klasifikasi banyak digunakan untuk memprediksi kelas pada suatu label tertentu, yaitu dengan mengklasifikasi data (membangun model) berdasarkan training set dan nilai-nilai (label kelas) dalam mengklasifikasikan atribut tertentu dan menggunakannya dalam mengklasifikasikan data yang baru [10].

Tujuan dari klasifikasi adalah untuk memprediksi secara akurat kelas/target untuk setiap kasus dalam data. Terdapat beberapa algoritma dalam klasifikasi, diantaranya, Bayesian Classification, Decision tree classification, Nearest Neighbor, Neural Networks(Back Propagation), Support Vector Machines (SVMs) [11].

2.2. Naïve Bayes Classifier

NBC adalah algoritma dalam pengklasifikasian statistik yang berguna untuk memprediksi probabilitas anggota dalam suatu kelas. NBC terbukti mempunyai akurasi yang tinggi apabila diterapkan ke dalam sebuah dataset yang besar [12]. Pengklasifikasian NBC diasumsikan bahwa ada atau tidak ciri tertentu dari suatu class tidak berkaitan dengan ciri class lainnya [13]. NBC merupakan algoritma yang akan digunakan untuk mengklasifikasi data kesejahteraan rumah tangga di Kabupaten Siak, Riau. Persamaan dari teorema Bayes adalah :

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

Dimana,

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

P(H) : Probabilitas hipotesis H (prior probabilitas)

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

P(X) : Probabilitas X

2.3. K-Fold Cross Validation

Cross validation ialah suatu metode statistik untuk mengevaluasi dan membandingkan algoritma pembelajaran dengan cara memecah data menjadi dua bagian. Data yang digunakan dalam memuat model disebut data latih, sedangkan data yang digunakan untuk memvalidasi disebut data uji [14].

Teknik ini utamanya digunakan untuk melakukan prediksi model dan memperkirakan seberapa akurat sebuah model prediktif ketika dijalankan dalam praktiknya. Salah satu teknik dari validasi silang adalah k-fold cross validation, yang mana memecah data menjadi k bagian set data dengan ukuran yang sama. Penggunaan k-fold cross validation untuk menghilangkan bias pada data. Tujuannya agar hasil prediksi atau klasifikasi tidak hanya akurasi tinggi melainkan juga valid [15].

2.4. Desil

Desil Merupakan Kelompok Rumah tangga dalam Basis Data Terpadu untuk Program Perlindungan Sosial yang dikelola oleh TNP2K. Desil adalah kelompok per-sepuluhannya sehingga seluruh rumah tangga dapat dibagi ke dalam 10 Desil. Kemudian Desil disesuaikan dengan keadaan kesejahteraan masing-masing daerah di Indonesia. Pengelompokan rumah tangga dalam Basis Data Terpadu adalah sebagai berikut (TNP2K, 2016):

- 1) Desil 1 adalah rumah tangga dalam kelompok 10% terendah (sangat miskin)
- 2) Desil 2 adalah rumah tangga dalam kelompok antara 11 sampai 20% terendah (miskin)
- 3) Desil 3 adalah rumah tangga dalam kelompok antara 21 sampai 30% terendah (hampir miskin)
- 4) Desil 4 adalah rumah tangga dalam kelompok antara 31 sampai 40% terendah (cukup miskin) dan seterusnya
- 5) Desil 10 adalah rumah tangga dalam kelompok 10% dengan tingkat kesejahteraan paling tinggi.

Basis Data Terpadu Kabupaten Siak berisikan kelompok Desil 1, Desil 2, dan Desil 3 serta Desil 4 karena memuat 40% rumah tangga dengan peringkat kesejahteraan terendah di Indonesia.

3. Hasil dan Analisis

3.1. Pengumpulan Data

Untuk melakukan penelitian ini, langkah pertama yang dilakukan adalah mengumpulkan data yang akan digunakan. Adapun tahapan pengumpulan data yaitu melakukan observasi dengan melihat masalah yang terjadi pada penentuan kesejahteraan keluarga dan melakukan pengambilan data sekunder serta melakukan wawancara, dimana wawancara dilakukan dengan pihak Dinas Sosial Kabupaten Siak terkait penentuan atribut yang akan digunakan.

Adapun atribut yang digunakan untuk klasifikasi data adalah NIK, Kecamatan, Jenis Cacat, Penyakit Kronis, Ijazah Tertinggi, Lapangan Usaha dari Pekerjaan Utama, Status Kedudukan dalam Pekerjaan Utama, Status Bangunan Tempat Tinggal, Status Lahan Tempat Tinggal, Jenis Lantai, Jenis Dinding, Jenis Atap, Sumber Air Minum, Sumber Penerangan Utama, dan Jumlah Anggota Keluarga.

Jumlah dataset yang digunakan dalam penelitian merupakan data kondisi tingkat kesejahteraan rumah tangga di Kabupaten Siak, Riau tahun 2016 yang di ambil dari Dinas Sosial Kabupaten Siak. Data yang digunakan dalam penelitian ini sebanyak 19.612 data yang dibagi menjadi data training dan data testing.

3.2. Pre-processing

Pada tahap ini, setelah mendapatkan data sekunder, selanjutnya yaitu melakukan proses cleaning. Data cleaning adalah proses menghilangkan data yang tidak relevan dan noise data. Biasanya data yang didapat dari database memiliki record yang tidak sempurna, seperti data hilang, data tidak valid atau salah ketik. Data yang seperti itu sebaiknya dibuang. Cleaning dapat mempengaruhi performa dari teknik data mining karena data yang diproses akan berkurang jumlah dan kompleksitasnya (Asriningtias dan Mardiyah, 2014). Pada tahap ini, awalnya data berjumlah 22.047 record data. Dan setelah dilakukan proses cleaning, maka jumlah data berubah menjadi 19.612 record data. Berikut ini adalah tampilan data setelah dicleaning terlihat pada tabel 1 dibawah ini.

Tabel 1. Data Cleaning

No.	NIK	Kecamatan	Penyakit Kronis	Ijazah Tertinggi	Lapangan Pekerjaan Utama	...	Status
1	12220xxx	Sungai Apit	Tidak ada	SD/ Sederajat	Perkebunan	...	Miskin
2	14080xxx	Sungai Apit	Asma	SD/ Sederajat	Lainnya	...	Sangat Miskin
3	14081xxx	Sabak Auh	Hipertensi	Tidak Punya Ijazah	Bangunan	...	Hampir Miskin
4	14081xxx	Kerinci Kanan	Tidak ada	Tidak Punya Ijazah	Lainnya	...	Hampir Miskin
...
19.612	14080xxx	Sungai Apit	Tidak ada	SD/ Sederajat	Lainnya	...	Cukup Miskin

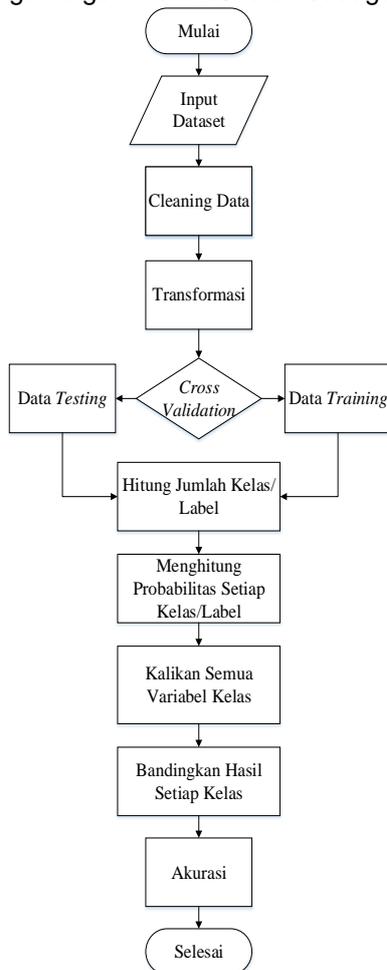
Setelah data melewati proses cleaning, selanjutnya ialah Transformasi data, transformasi data yaitu data diubah menjadi bentuk yang sesuai untuk diproses dalam data mining. Salah satu tujuan dari transformasi ini ialah agar data dapat diolah dengan menggunakan sebuah Tools. Berikut ini adalah hasil transformasi data terlihat pada tabel 2.

Tabel 2. Data Transformasi

No.	NIK	Kecamatan	Penyakit Kronis	Ijazah Tertinggi	Lapangan Pekerjaan Utama	...	Status
1	12220xxx	12	0	1	3	...	Miskin
2	14080xxx	12	3	1	21	...	Sangat Miskin
3	14081xxx	10	1	0	11	...	Hampir Miskin
4	14081xxx	4	0	0	21	...	Hampir Miskin
...
19.612	14080xxx	12	0	1	21	...	Cukup Miskin

3.3. Naïve Bayes Classifier

Adapun alur dari perhitungan algoritma NBC ialah sebagai berikut.

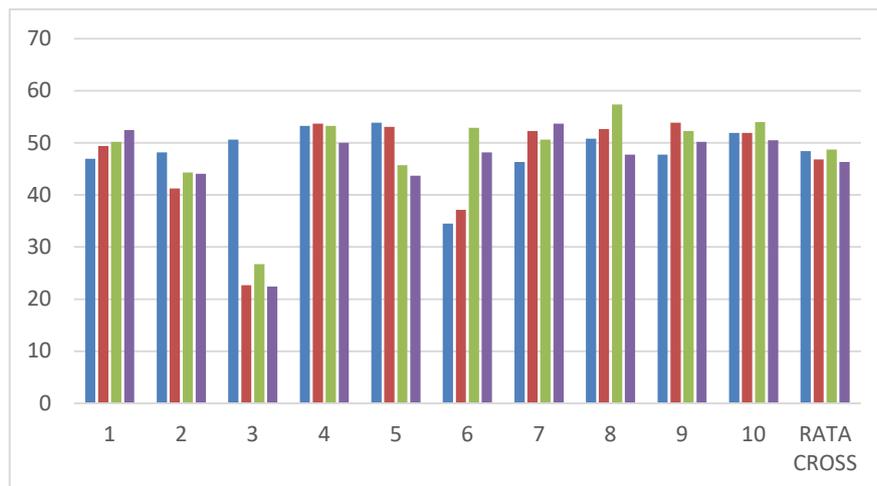


Gambar 2. Alur NBC

Algoritma ini mengasumsikan bahwa semua atribut bersifat independen atau tidak saling berketergantungan [16][16]. Pada tahapan ini, K-Fold Cross Validation digunakan untuk menentukan parameter terbaik dalam suatu model ini akan dilakukan dengan melakukan uji besarnya akurasi pada data testing. Model dan parameter “K” data sampel yang digunakan adalah nilai akurasi paling tinggi dan nilai error paling rendah. Dikarenakan jumlah data record sebanyak 19.612 maka dalam cross validation akan dibagi secara merata sebanyak 4 model bagian maka secara pembagian dapat dihitung sebagai berikut:

$$\text{Perbandingan} = \frac{19.612}{4}$$

Dalam cross validation data status kesejahteraan rumah tangga Kabupaten Siak akan diuji dalam 4 sampel. Dalam 3-1 sampel sebagai data training, dan 1 sampel sebagai data testing. Permodelan cross validation dapat di lihat pada gambar 3. Cross validation mengambil 10 uji untuk menentukan nilai parameter “K” dan model yang baik untuk digunakan pada NBC. Cross validation di Uji menggunakan Tool / Aplikasi Weka Machine Learning. Berikut adalah grafik hasil dari pengukuran akurasi dengan pengujian untuk menentukan nilai 10 nilai parameter K dengan K-Fold Cross Validation dapat dilihat pada gambar 3.



Gambar 3. Hasil pengukuran akurasi

Dari Gambar 3, terlihat bahwa nilai parameter K tertinggi berada pada K4 dengan rata-rata akurasi diatas 50. Kemudian dengan nilai parameter K terendah berada pada K3 dengan rata-rata akurasi 30. Hasil pengukuran akurasi secara rinci, dapat dilihat pada Tabel 3.

Tabel 4. Data hasil pengukuran akurasi

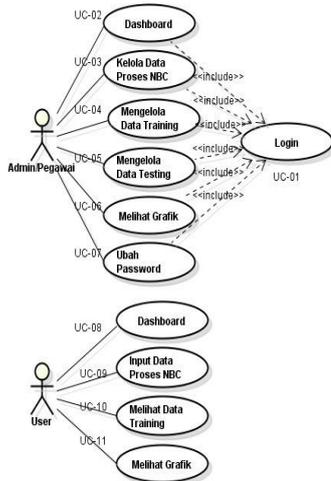
K	Cross 1	Cross 2	Cross 3	Cross 4	Rata K
1	46,9388	49,3878	50,2041	52,4490	49,7449
2	48,1633	41,2245	44,2857	44,0816	44,4388
3	50,6122	22,6531	26,7347	22,4490	30,6123
4	53,2653	53,6735	53,2653	50,0000	52,5510
5	53,8776	53,0612	45,7143	43,6735	49,0817
6	34,4898	37,1429	52,8571	48,1633	43,1633
7	46,3265	52,2449	50,6122	53,6735	50,7143
8	50,8163	52,6531	57,3469	47,7551	52,1429
9	47,7551	53,8776	52,2449	50,2041	51,0204
10	51,9270	51,9270	53,9554	50,5071	52,0791
Rata Cross	48,4172	46,7846	48,7221	46,2956	-

Berdasarkan hasil dari cross validation dapat dilihat bahwa parameter “K” terbaik terdapat pada K=4 dengan rata-rata akurasi adalah 52,5510% dan model “cross” yang baik terdapat pada cross 3 dengan rata-rata akurasi adalah 48,7221%. Selanjutnya jumlah data training pada cross 3 akan digunakan sebagai data training pada Sistem NBC dengan menggunakan K=4.

3.4. Analisis Kebutuhan Sistem

Pada tahap ini, ditentukan fungsionalitas apa saja yang dibutuhkan oleh sistem informasi yang akan dibangun untuk menunjang dan mendukung kinerja operasional. Pada tahap ini akan dilakukan rancangan sistem klasifikasi untuk menentukan status kesejahteraan rumah tangga di Kabupaten Siak, menggunakan pendekatan berorientasi objek yaitu dengan menggunakan diagram Unifield Modeling Language (UML). Adapun diagram UML yang digunakan terdiri atas tiga bagian/diagram, yaitu:

a) Usecase Diagram



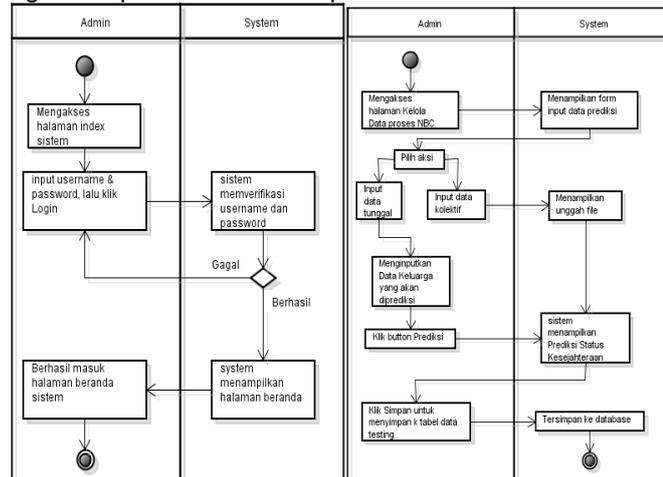
b) Class Diagram



Gambar 4. Usecase diagram

Gambar 5. Class diagram

c) Activity Diagram login dan proses klasifikasi pada sistem

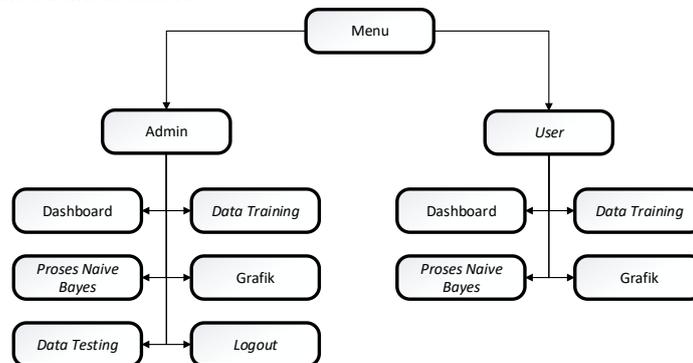


Gambar 6. Activity diagram

3.5. Perancangan Sistem

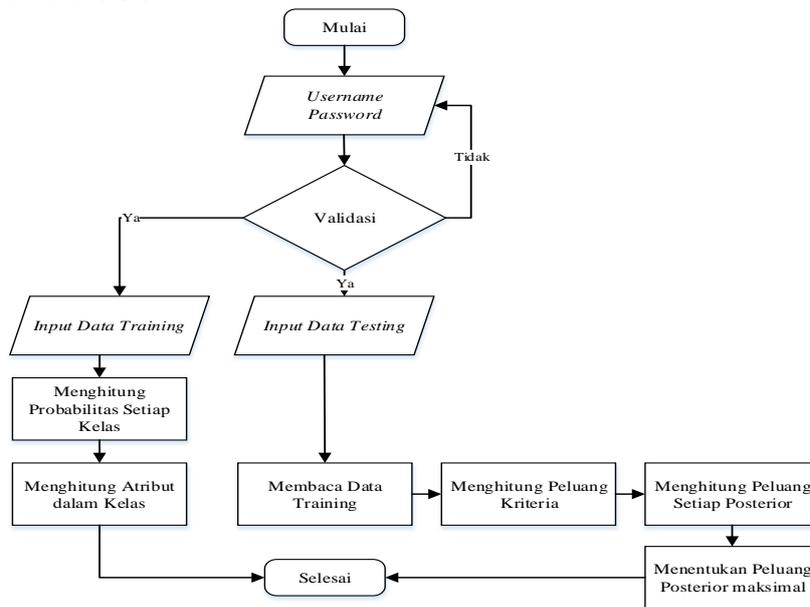
Perancangan sistem perlu dilakukan sebelum dilakukan pembuatan sistem. Tujuan dari perancangan sistem adalah untuk menentukan, mengorganisir, dan membentuk komponen dari solusi sistem akhir sehingga memiliki blueprint untuk membangun sistem.

a) Perancangan struktur menu



Gambar 7. Struktur menu

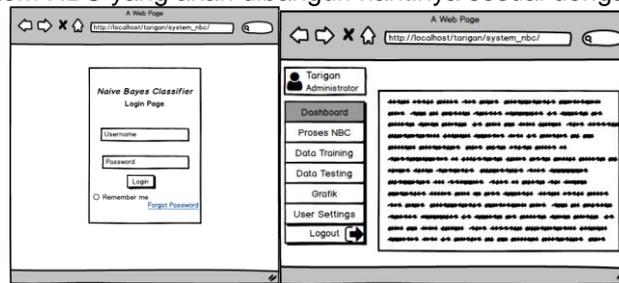
b) Flowchart alur sistem



Gambar 8. Flowchart sistem

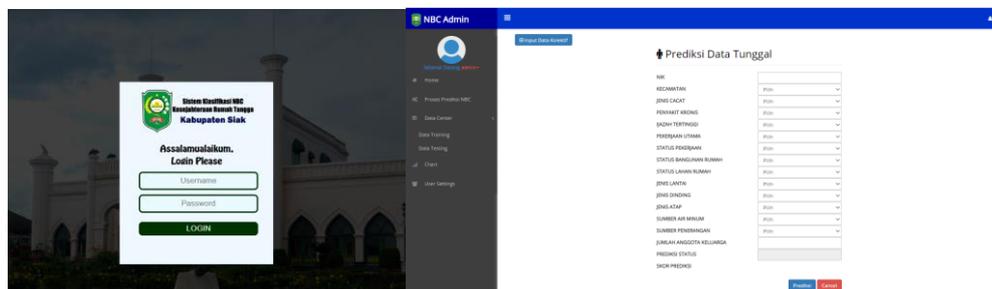
c) Perancangan Infrface

Rancangan interface system dibawah ini merupakan gambaran beberapa rancangan antarmuka sistem NBC yang akan dibangun nantinya sesuai dengan kebutuhan pengguna.



Gambar 9. Rancangan interface sistem

3.6. Implementasi Sistem



Gambar 10. Halaman Admin

3.7. Pengujian Blackbox Sistem

Pengujian merupakan tahapan untuk mengetahui suatu program dengan tujuan menemukan suatu kesalahan. Metode black box testing berfokus pada spesifikasi fungsional dari perangkat lunak. Tester dapat mendefinisikan kumpulan kondisi input dan melakukan pengujian pada spesifikasi fungsional program [11]. Hasil pengujian dapat dilihat pada Tabel 4.

Tabel 4. Pengujian Blackbox

No	Deskripsi Pengujian	Hasil Yang Diharapkan	Kriteria Evaluasi Hasil	
			Berhasil	Tidak Berhasil
1	Modul <i>Login</i>	Masuk ke halaman utama admin	√	
2	Modul <i>dashboard</i>	Masuk ke halaman utama sistem	√	
3	Modul Data <i>Training</i>	Data yang dipilih muncul dan data training yang diolah tersimpan di dalam <i>database</i>	√	
4	Modul Proses Klasifikasi NBC	Diperoleh hasil perhitungan dan Prediksi Kelas	√	
5	Modul Data testing	Data yang diuji muncul dan data testing yang diolah tersimpan di dalam <i>database</i>	√	
6	Modul Lihat Grafik	Informasi jumlah kelas perkecamatan dalam bentuk grafik.	√	

4. Kesimpulan

- a) Algoritma NBC dalam klasifikasi status kesejahteraan rumah tangga Kabupaten Siak menghasilkan akurasi dengan nilai rata-rata 53%. Pengujian algoritma ini dilakukan menggunakan perhitungan confusion matriks yaitu membandingkan kelas label faktual dengan kelas label klasifikasi menggunakan tools weka.
- b) Proses klasifikasi status kesejahteraan rumah tangga Kabupaten Siak dengan membangun Sistem Informasi berdasarkan algoritma Naive Bayes Classifier (NBC). Sistem yang dibangun mampu mengklasifikasi sesuai dengan hasil perhitungan Algoritma secara manual. Data yang digunakan adalah 19.612 record. Data training 14.709 dan 4.903 record data testing dari data Kesejahteraan tahun 2017.

5. Saran

- a) Proses analisa yang digunakan dalam proses melakukan klasifikasi menentukan status kesejahteraan rumah tangga menggunakan algoritma Naive Bayes Classifier menunjukkan akurasi yang cukup rendah, hal ini menunjukkan bahwa algoritma yang digunakan tidak sesuai untuk data pada kasus tersebut dan untuk selanjutnya agar dapat dilakukan dengan beberapa algoritma lainnya seperti Decision Tree, Support Vector Machine (SVM), Probabilistic Neural Network (PNN) dan algoritma klasifikasi lainnya untuk mengetahui akurasi terbaik dari beberapa algoritma pada kasus yang sama.
- b) Sistem Klasifikasi penentu status Kesejahteraan rumah tangga Kabupaten Siak perlu dikembangkan lagi untuk meningkatkan kualitas dan performa sistem sehingga interaksi antara sistem dan user lebih baik.

Daftar Pustaka

- [1] E. Sunarti, "Indikator Keluarga Sejahtera," 2006.
- [2] Riska Prakasita Sahitayakti dan Kartika Fithriasar, "Klasifikasi Kesejahteraan Rumah Tangga di Provinsi Papua dengan Metode Regresi Logistik dan Support Vector Machine," vol. 4, no. 2, pp. 2–8, 2015.
- [3] J. Maulani, "Klasifikasi Kesejahteraan Rumah Tangga dengan Penerapan Algoritma Decision Tree Seleksi Atribut Genetik Algoritma," *Technologia*, vol. 7, no. 3, pp. 137–141, 2016.
- [4] T. Marnoto, "Comparative Analysis of Data Mining Classification Algorithms in Type-2 Diabetes Prediction Data Using WEKA Approach," *Int. J. Sci. Eng.*, vol. 7, no. 2, pp. 155–160, 2014.
- [5] F. J. Simatupang, T. Wuryandari, and Suparti, "Klasifikasi Rumah Layak Huni Di Kabupaten Brebes Dengan Menggunakan Metode Learning Vector Quantization Dan Naive Bayes," *Gaussian*, vol. 5, no. 1, pp. 99–111, 2016.
- [6] D. W. Nugraha, A. Y. E. Dodu, and N. Chandra, "Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes Classifier (Studi Kasus Pada Rumah Sakit Umum Daerah Undata Palu)," vol. 3, no. 2, pp. 13–22, 2017.
- [7] E. Karyadiputra, S. Kom, and M. Kom, "Analisis Algoritma Naive Bayes Untuk Klasifikasi Status Kesejahteraan Rumah Tangga Keluarga Binaan Sosial," vol. 7, no. 4, pp. 199–208, 2016.
- [8] M. Paramita and R. Ely, "Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma

- Klasifikasi Data Mining,” *Semin. Nas. Sist. Inf. Indones.* 2015, vol. 11, no. November, pp. 1–7, 2015.
- [9] Y. F. Safri, R. Arifudin, and M. A. Muslim, “K-Nearest Neighbor and Naive Bayes Classifier Algorithm in Determining The Classification of Healthy Card Indonesia Giving to The Poor,” *Sci. J. Informatics*, vol. 5, no. 1, p. 18, 2018.
- [10] V. Jain, G. S. Narula, and M. Singh, “Implementation of Data Mining in Online Shopping System Using,” vol. 2, no. 1, pp. 47–58, 2013.
- [11] P. G. Espejo, S. Ventura, and F. Herrera, “A survey on the application of genetic programming to classification,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 40, no. 2, pp. 121–144, 2011.
- [12] W. B. Zulfikar and N. Lukman, “Perbandingan Naive Bayes Classifier Dengan Nearest Neighbor Untuk Identifikasi Penyakit Mata,” *J. Online Inform.*, vol. 1, no. 2, pp. 82–86, 2016.
- [13] V. Agarwal, S. Thakare, and A. Jaiswal, “Survey on Classification Techniques for Data Mining,” *Int. J. Comput. Appl.*, vol. 132, no. 4, pp. 13–16, 2015.
- [14] Refaeilzadeh, “Cross-Validation,” *Arizona State Univ.*, 2008.
- [15] F. Tempola, M. Muhammad, and A. Khairan, “Perbandingan Klasifikasi Antara Knn Dan Naive Bayes Pada Penentuan Status Gunung Berapi Dengan K-Fold Cross Validation Comparison of Classification Between Knn and Naive Bayes At the Determination of the Volcanic Status With K-Fold Cross,” vol. 5, no. 5, pp. 577–584, 2018.
- [16] T. . Patil and S. . Shrekar, “Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification,” *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, no. 2, pp. 88–91, 2019.
- [17] Hidayat, A., Effendi, Z., Aszani., Novita, R., Lestari, T.L. “Algorithm Comparison Of Naive Bayes Classifier And Probabilistic Neural Network For Water Area Classification Of Fishing Vessel In Indonesia”. *Journal of Theoretical and Applied Information Technology*, Vol. 96 No. 13 Hal 4114-4125, Juli 2018.