

PENERAPAN ALGORITMA GENETIKA DALAM MEMREDIKSI PENERIMA PROGRAM KELUARGA HARAPAN DENGAN METODE NAÏVE BAYES

Wahyu Hadikristanto¹⁾, Imam Nasai²⁾

Program Studi Teknik Informatika Fakultas Teknik
Universitas Pelita Bangsa
wahyu.hadikristanto@pelitabangsa.ac.id

Disetujui 25 September 2019

Abstraksi

Program keluarga harapan adalah sebuah program pemerintah yang bertujuan untuk meringankan beban keluarga miskin atau hampir miskin dalam hal pangan, dalam implementasinya masih belum optimal dikarenakan masih banyak program keluarga harapan yang belum tepat sasaran, untuk menghindari adanya penentuan penerimaan secara subjektif maka konsep data mining akan mempermudah mengatasi masalah yang belum optimal maka metode klasifikasi mampu menemukan tujuan pemilihan penerimaan bantuan. Penerapan genetika dapat memprediksi peluang bertujaun untuk menentukan perbedaan hasilnya. Pada penelitian ini mengambil data yang lati sebanyak 500 data. Metode klasifikasi yang digunakan dalam penelitian ini adalah Algoritma Genetik dan Naive Bayes. Hasil penguji PKH untuk metode Naive Bayes dalam accuracy sebanyak 94,80% dan hasil algoritma genetic 69,00% maka perbandingan tersebut turun sebanyak 25,80%. Hasil penguji PKH untuk metode Naive Bayes dalam presicion sebanyak 94,48% dan hasil algoritma genetic 0,00% maka perbandingan tersebut tetap tidak ada perubahan. Hasil penguji PKH untuk metode Naive Bayes dalam recall sebanyak 88,38% dan hasil algoritma genetic 0,00% maka perbandingan tersebut tetap tidak ada perubahan. Dengan demikian klasifikasi menggunakan metode naive bayes classification dapat digunakan untuk pemilihan program bantuan terharap keluarga kurang mampu.

Kata Kunci : Naive Bayes, Algoritma Genetik, RapidMiner, PKH

Abstract

Hope family program is a government program that aims to relieve the burden of poor or near poor families in terms of food, in its implementation is still not optimal because there are still many hopeful family programs that are not yet on target, to avoid any subjective determination of the data mining concept will facilitate overcome problems that have not been optimal, the classification method is able to find the purpose of selecting aid acceptance. The application of genetika can predict the opportunity to visit to determine the difference in results. In this study took 500 data lati data. The lassification methods used in this study are Genetic Algorithms and Naive Bayes. The results of the layoff test for the Naive Bayes method in accuracy were 94.80% and the genetic algorithm results 69.00% then the ratio dropped by 25.80%. The results of the layoff test for the Naive Bayes method in the percentage were 94.48% and the genetic algorithm results were 0.00% then the comparison remained unchanged. The results of layoff test for the Naive Bayes method in recall were 88.38% and the genetic algorithm results were 0.00% then the comparison remained unchanged. Thus the classification using the naive bayes classification method can be used for selecting aid programs for underprivileged families.

Keywords : Naive Bayes, Algoritma Genetic, RapidMiner, PKH

1. Pendahuluan

Indonesia merupakan negara yang mempunyai lebih dari 269.000.000 penduduk yang tersebar di beberapa provinsi berdasarkan data yang dihimpun oleh Badan Pusat Statistik tahun 2019. Selain memiliki jumlah penduduk yang cukup besar tersebut, Indonesia memilki beragam masalah social antara lain tingkat kejahatan yang tinggi, swasembada pangan, masalah ekonomi dan lain lain.

Kemiskinan merupakan suatu masalah yang masih terjadi di negara berkembang, termasuk negara Indonesia. Salah satu kebijakan untuk mengatasi tingkat kemiskinan adalah dengan mengadakan Program Keluarga Harapan (PKH). Tingkat kemiskinan di Indonesia mengalami penurunan [1]. Program Keluarga Harapan (PKH), merupakan salah satu program penanggulangan kemiskinan kedudukan PKH merupakan bagian dari program – program penanggulangan kemiskinan layanan.

Program Keluarga Harapan (PKH) ada sejak tahun 2007 yaitu dengan memberikan Bantuan Tunai Bersyarat (BTB) yang dikenal dengan nama Program Keluarga Harapan (PKH) sebagai salah satu tahapan menuju sistem perlindungan sosial. Program perlindungan sosial yang juga dikenal di dunia Internasional dengan istilah Conditional Cash Transfer (CCT) ini terbukti cukup berhasil dalam menanggulangi kemiskinan yang dihadapi di Negaranegara tersebut, terutama masalah kemiskinan kronis (Kemensos RI, 2017: 12).

Program Keluarga Harapan adalah suatu program yang memberikan bantuan tunai kepada Rumah Tangga Sangat Miskin (RTSM), jika mereka memenuhi persyaratan yang terkait dengan upaya peningkatan kualitas sumberdaya manusia (SDM), yaitu pendidikan dan kesehatan. Umumnya pemilihan penerima PKH di desa atau dilakukan oleh kepala dusun. Untuk menghindari adanya penentuan penerima PKH secara subyektif, maka diperlukan sebuah metode yang mampu mengatasi hal tersebut. Hal ini diharapkan dapat membantu pihak desa dalam pengambilan keputusan penerima PKH. Pelaksanaan Keluarga Harapan tersebut khusus nya sudah lama terlaksana sejak di sahkan kebijakan bantuan tersebut oleh pemerintah. Namun berdasarkan data terakhir, RTSM peserta PKH masih merupakan persentase yang kecil dari jumlah RTSM yang tersebar, artinya negara belum mampu menjalankan fungsi perlindungan social secara optimal. Berdasarkan persoalan tersebut penelitian ini akan membahas dan membuat pendukung keputusan yang diharapkan dapat membantu penyeleksi dalam pemilihan calon penerima PKH, yaitu berupa kelayakan calon penerima dan sebagai rekomendasi dari pengambilan keputusan untuk memilih keluarga yang tepat untuk menerima PKH. Setiap warga mempunyai kriteria nilai yang berbeda beda terhadap penentuan syarat dari penyeleksi dan penentuan calon keluarganya yang layak dan tepat. Oleh sebab itu maka diperlukan sebuah alat bantu yang dapat digunakan untuk melakukan seleksi terhadap calon penerima PKH dari RTSM yang terdata.

2. Tinjauan Studi

Dalam sebuah penelitian terutama pada saat menganalisis sistem, akan lebih mudah jika menggunakan metode pendekatan dan pengembangan sistem.

2.1. Data Mining

Menurut Han dan Kamber (2015,p36), data mining adalah proses menemukan pola yang menarik dan pengetahuan dari data yang berjumlah besar. Sedangkan menurut Linoff dan Berry (2017, p7) Data mining adalah suatu pencarian dan analisa dari jumlah data yang sangat besar dan bertujuan untuk mencari arti dari pola dan aturan. Lalu menurut Connolly dan Begg,(2013), Data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahui sebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting Metode Pengembangan Sistem

2.2. Klasifikasi

Menurut Olson dan Shi (2016),Klasifikasi (Classification),metode-metodenya ditunjukkan untuk pembelajaran fungsi-fungsi berbeda yang memetakan masing-masing data terpilih ke dalam salah satu dari kelompok kelas yang telah ditetapkan sebelumnya. Menurut Han dan Kamber (2011, 327), Classification adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data,dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dasar pengukuran untuk mengukur kualitas dari penemuan teks fitur kalkulasi dan pembuatan grafik yang berupa pengolah angka.

2.3. Metode Naïve Bayes

Naïve Bayes Classifier (NBC) adalah metode *classifier* berdasarkan probabilitas dan *teorema bayesian* dengan asumsi bahwa setiap variabel X bersifat bebas (*indepedence*) (Buntoro,2017),NBC merupakan metode klasifikasi statistik untuk memprediksi probabilitas suatu atribut dalam suatu kelas, dalam proses klasifikasinya menganggap bahwa kemunculan suatu kata pada suatu kalimat tidak dipengaruhi oleh kemunculan kata-kata yang lain pada kalimat tersebut,namun pada kenyataannya kemunculan suatu kata dipengaruhi oleh kata-kata lainnya, atau disebut dengan *naïve*(Susanti,2016).

2.4. RapidMiner

RapidMiner berkembang sejak tahun 2018, sebelumnya disebut dengan nama YALE (Yet Another

Learning Environment). Software ini dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, serta Simon Fischer pada Unit Artificial Intelligence dari Technical University of Dortmund. RapidMiner adalah platform analisis modern yang meliputi data mining, machine learning, analisis prediktif, text mining dan analisis bisnis.

2.5. Algoritma Genetika

Algoritma Genetika adalah proses pencarian yang mengikuti proses evolusi. Dalam proses ini akan dicari solusi yang terbaik dari yang terbaik dari pembangkitan sebelumnya sehingga didapatkan solusi yang paling optimal [29]. Beberapa hal yang harus dilakukan dalam Algoritma Genetika adalah :

- 1) Mendefinisikan individu, definisi individu disini adalah solusi dari permasalahan yang ada.
- 2) Mendefinisikan nilai fitness, nilai fitness merupakan ukuran baik tidaknya sebuah solusi yang didapatkan.
- 3) Menentukan proses pembangkitan populasi awal
- 4) Menentukan proses seleksi yang akan digunakan.
- 5) Menentukan proses perkawinan silang (crossover) dan mutasi gen.

3. Kerangka Konsep

Objek Penelitian

Dalam penyusunan penelitian ini, lingkup objek penelitian di tetapkan peneliti sesuai dengan permasalahan yang akan diteliti adalah menentukan penerimaan bantuan program keluarga harapan (PKH) tetap sasaran supaya untuk menghindari adanya penentuan penerima PKH secara subyektif, maka diperlukan sebuah metode yang mampu mengatasi hal tersebut.



Gambar 1. Metode Penelitian

1. Pengumpulan Data

Pada tahapan ini dijelaskan tentang bagaimana dan dari mana data dalam penelitian ini didapatkan, meliputi data sekunder dan data primer. Data sekunder berisi tentang sumber perolehan data untuk kepentingan penelitian, sedangkan data primer berisi tentang data yang dihasilkan dari penelitian.

2. Pengolahan Data Awal

Pada tahapan ini di jelaskan tentang tahap awal data mining. Pengolahan awal data meliputi proses input data ke format yang dibutuhkan, penentuan dan pengelompokan atribut data.

3. Metode Yang Diusulkan

Pada tahapan ini dijelaskan tentang metode yang diusulkan untuk digunakan pada prediksi ujian kompetensi. Penjelasan meliputi pemilihan dan pengaturan nilai dari parameter-parameter dan arsitektur melalui uji coba.

4. Eksperimen dan Pengujian Metode

Pada tahapan ini dijelaskan tahapan-tahapan eksperimen meliputi cara pemilihan arsitektur yang tepat dari metode atau model yang diusulkan sehingga didapatkan hasil yang dapat membuktikan bahwa metode yang digunakan adalah tepat.

5. Evaluasi dan Validasi Hasil

Pada tahapan ini dijelaskan tentang validasi dan evaluasi hasil penerapan metode dan pengujian Algoritma Genetika pada penelitian yang dilakukan.

4. Desain Penelitian/Methodologi

Dalam penelitian ini membuat sebuah pemikiran yang berguna sebagai pedoman atau acuan dalam

penelitian ini, sehingga dapat dilakukan secara konsisten. Metode yang digunakan dalam penelitian ini yaitu dengan menggunakan metode naïve bayes untuk memecahkan sebuah masalah yang berkaitan dengan penelitian, menemukan kriteria keluarga dalam menentukan bantuan dalam program keluarga harapan (PKH).



Gambar 2. Kerangka Pemikiran

4.1 Pengumpulan Data

Pengalaman pakar yang di jadikan sumber antara lain jurnal ilmiah. Berbasis pengetahuan menggunakan hasil diskusi dari beberapa pengetahuan dari jurnal ilmiah. Dalam metode pengumpulan data terdapat 2 metode yaitu pengumpulan data primer dan sekunder :

1. Data primer adalah data yang dikumpulkan melalui pihak pertama, biasanya dapat melalui wawancara, jurnal ilmiah dan lain-lain.
2. Data sekunder adalah data yang dikumpulkan dari tangan kedua atau dari sumber-sumber lain yang telah tersedia sebelum penelitian dilakukan.

Tabel 1. Contoh Data Set

No	Nama Kepala Rumah	PKH	Jumlah Tanggungan	Kepala Rumah tangga	Rumah	Jumlah Penghasilan	Kondisi Rumah	Keterangan
1	MUNIDEH	Tidak	2	Laki - Laki	Batu	Tinggi	Layak	idak Miski
2	MARYAM	Tidak	4	Laki - Laki	Bambu	Rendah	Layak	idak Miski
3	JULIA	Tidak	1	Laki - Laki	Batu	Tinggi	idak Laya	Miskin
4	SANINTI	Tidak	4	Laki - Laki	Batu	Tinggi	Layak	idak Miski
5	SUTARI	Tidak	3	Laki - Laki	Papan	Sedang	Layak	idak Miski
6	ASNI	Tidak	1	Laki - Laki	Batu	Tinggi	idak Laya	Miskin
7	SULIYAH	Tidak	1	Laki - Laki	Bambu	Rendah	Layak	idak Miski
8	RIDO A	Tidak	5	Laki - Laki	Bambu	Tinggi	Layak	idak Miski

9	ISNA YUNE	butuh	2	Laki - Laki	Papan	Rendah	idak Laya	Miskin
10	TUBEH	Tidak	4	Laki - Laki	Batu	Tinggi	idak Laya	Miskin
11	HORIAH	Tidak	1	Laki - Laki	Papan	Sedang	idak Laya	Miskin
12	PUAH	Tidak	2	Laki - Laki	Bambu	Rendah	Layak	idak Miski
13	SATIYEH	Tidak	3	Perempuua	Bambu	Rendah	Layak	idak Miski
14	YANTI	Tidak	4	Laki - Laki	Bambu	Rendah	Layak	idak Miski
15	ZAINAB	Tidak	2	Perempuan	Papan	Sedang	Layak	idak Miski
16	SUNI	butuh	1	Laki - Laki	Batu	Tinggi	Layak	idak Miski
17	IDAYANI	Tidak	4	Laki - Laki	Batu	Tinggi	Layak	idak Miski
18	HAMIDAH	Tidak	1	Laki - Laki	Batu	Tinggi	idak Laya	Miskin
19	SULAN	Tidak	2	Laki - Laki	Batu	Tinggi	idak Laya	Miskin
20	MAR/SIDE	Tidak	2	Laki - Laki	Batu	Tinggi	Layak	idak Miski
21	SUHAMA	butuh	4	Laki - Laki	Bambu	Rendah	Layak	idak Miski
22	NURYA	Tidak	2	Laki - Laki	Bambu	Rendah	Layak	idak Miski
23	NASIYEH	Tidak	1	Laki - Laki	Papan	Sedang	idak Laya	Miskin
24	NURAI SYAH	butuh	3	Laki - Laki	Batu	Tinggi	Layak	idak Miski
25	MAIMUNAH	Tidak	4	Laki - Laki	Batu	Tinggi	idak Laya	Miskin
26	SIMAH	butuh	2	Laki - Laki	Batu	Tinggi	idak Laya	Miskin
27	MISNAWAT	Tidak	1	Laki - Laki	Papan	Sedang	idak Laya	Miskin
28	B. ROBIH	butuh	2	Laki - Laki	Bambu	Rendah	Layak	idak Miski
29	SUHARNI	butuh	4	Laki - Laki	Bambu	Rendah	Layak	idak Miski
30	SAYUTI	butuh	1	Laki - Laki	Batu	Tinggi	Layak	idak Miski

Untuk menentukan data yang nantinya akan dianalisis dengan metode naïve bayes maka langkah pertama yang akan dilakukan adalah membaca data latih, adapun data latih yang dapat di lihat pada table berikut.

4.2 Data Atribut yang di gunakan

Pada data nama kepala keluarga yang telah dikumpulkan dan dilakukan *cleaning* data untuk menghilangkan *record-record* yang *noise* atau tidak lengkap atau *record* yang berulang. Atribut yang tidak diperlukan seperti atribut nama. Hal ini dilakukan karena atribut tersebut tidak berpengaruh terhadap

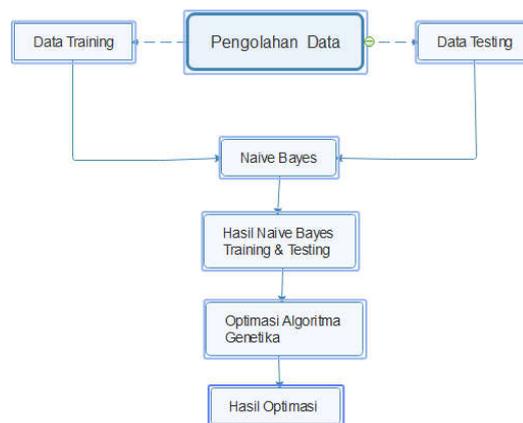
pengolahan data pada proses selanjutnya. Sedangkan untuk atribut seperti status, nama kepala keluarga, pkh, jumlah tanggungan, rumah, kondisi rumah Hal ini dilakukan karena atribut tersebut berpengaruh terhadap pengolahan data pada proses selanjutnya. Berikut adalah tabel dan katagorinya.

Tabel 2. Atribut Yang Digunakan

No	Atribut	Keterangan
1	Nama Kepala keluarga	Nama keluarga keluarga yang menerima PKH
2	PKH	Program pemberian uang tunai kepada rumah tangga sangat miskin berdasarkan persyaratan kriteria tertentu
3	Jumlah Tanggungan	Jumlah tanggungan keluarga
4	Rumah	Rumha yang terbuat dari, papan, bambu dan batu
5	Kondisi Rumah	Layak atau tidak untuk menempati rumah tersebut

4.3 Metode Yang Diusulkan

Dalam penelitian ini akan dilakukan analisa menggunakan metode *Naive bayes* dan optimasi Algoritma Genetika. Dalam penelitian ini untuk pengelohan data PHK dengan penggunaan *Naive Bayes*. Data dihitung dengan menggunakan algoritma sesuai dengan metodenya kemudian dibandingkan metode optimasi Algoritma Genetika dengan melihat perbandingan akurasi tertinggi. Dalam Tahapan ini akan dilakukan beberpa langkah pengujian data yaitu sebagai berikut :



Gambar 3. Tahap Pengujian Data

Data PHK akan diolah dengan dua metode, kemudian dilihat akurasinya pada setiap metode. Hasil pengujian dengan akurasi yang paling tinggi adalah metode yang akan digunakan untuk memprediksi pengguna PHK.

1. *Naive Bayes* yaitu salah satu algoritma dalam metode pohon keputusan yang merubah data
2. Algoritma Gentika yaitu salah satu metode penyelesaian optimasi yang dikenal mampu menghasilkan nilai optimum, dan juga memberi jalur optimum sesuai dengan yang diharapkan.

5. Hasil Penelitian dan Pembahasan

5.1 Proses Pengolahan Dataset

Pada penelitian ini dataset yang digunakan ialah data yang berasal dari <https://data.pontianakkota.go.id/km/dataset/program-keluarga-harapan>. Adapun tahapan proses pengolahannya hingga menjadi *dataset* yang nantinya akan digunakan untuk pengujian penelitian ini, berikut tahapan proses pengolahan dataset.

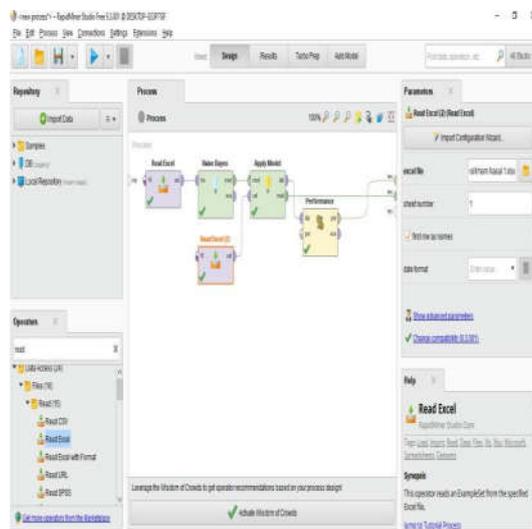
Dari data diatas merupakan data keseluruhan, data keseluruhan berjumlah 500 data penerima. Untuk melakukan pengolahan data di *RapidMiner* data yang digunaka adalah *Data Training*, dari seluruh data dipisah menjadi 80% untuk *data training* yang dipakai untuk membuat model atau *rule* dan 20%

sebagai *data testing* yang dipakai untuk sampel *testing* dan data dipilih secara acak. Besarnya data yang diperoleh adalah sebagai berikut yang ditunjukkan pada **Tabel 4.2**

Tabel 3. Hasil Pembagian Data

Data	500	
Training	80%	400
Testing	20%	

Dari data diatas merupakan *data training*, *data training* adalah 80% dari data keseluruhan, total keseluruhan data training berjumlah 400 data. Dari 400 *data training* dipilih secara acak. Atribut yang dipakai hanya atribut nama keluarga, PKH, kepala rumah tangga, kondisi rumah, dan jumlah penghasilan.

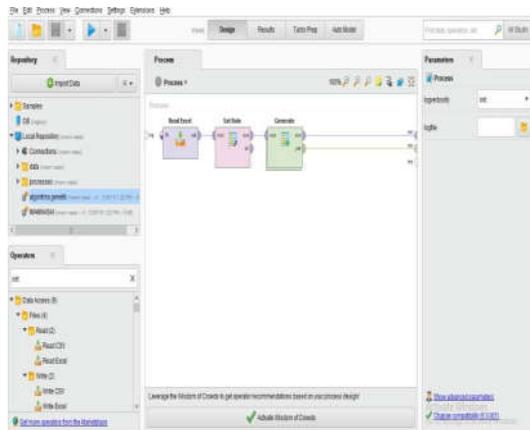


Gambar 4. Tampilan Algoritma Naïve Bayes Dari hasil pengujian yang telah dilakukan

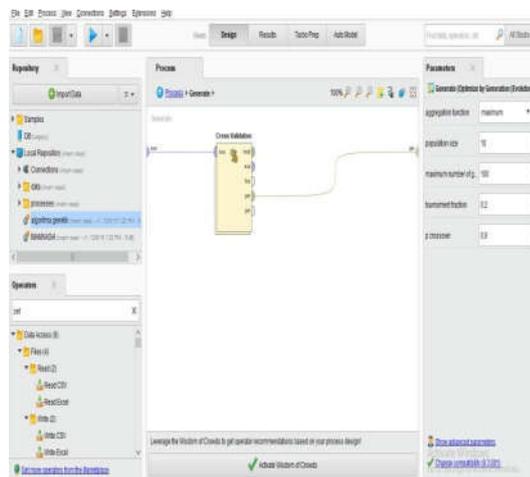
menggunakan algoritma *Naïve Bayes* dan dilakukanya pengujian menggunakan teknik *cross validation* peneliti mendapatkan hasil pengujian seperti pada tabel berikut:

Prediksi	Metode		Perbandingan	Keterangan
	Naïve Bayes	Naïve Bayes + GA		
Accuracy	94,80%	69,00%	25.80%	Turun
Precision	94,48%	0,00%	0,00%	Tetap
Recall	88,38%	0,00%	0,00%	Tetap

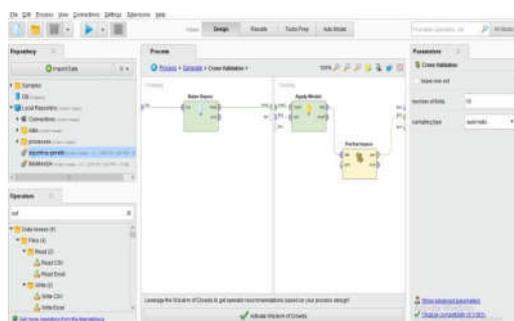
5.2 Pengujian Akurasi *Naïve Bayes* Optimasi Algoritma Genetika



Gambar 5. Tampilan Optimasi Algoritmat Genetika

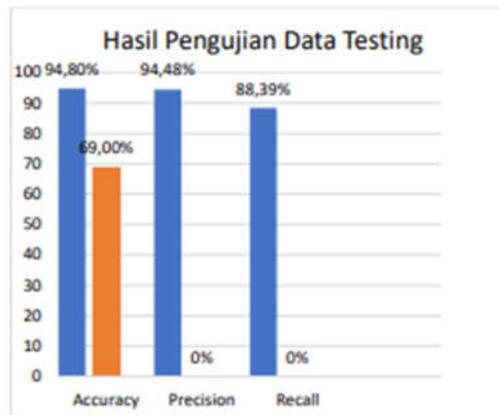


Gambar 6. Tampilan Sub Main Proses



Gambar 7. Tampilan Cross Validation

5.3 Pembahasan Hasil Pengujian Perbandingan



Tabel 4. Tabel Hasil Akurasi *Naive Bayes* & Algoritma Genetika

Prediksi	Metode		Perbandingan	Keterangan
	Naive Bayes	Naive Bayes + GA		
Accuracy	94,80%	69,00%	25,80%	Turun
Precision	94,48%	0,00%	0,00%	Tetap
Recall	88,38%	0,00%	0,00%	Tetap

6 Kesimpulan

Dari penerapan metode yang diterapkan yaitu *Naive Bayes* dapat di ketahui bahwa metode *Naive Bayes* lebih baik kinerjanya dalam penelitian penentuan program PKH dari sejumlah data 500 calon penerima program PKH dan bias menentukan butuh dan tidak butuh menerima bantuan program PKH tersebut. Dan hasil yang diperoleh dari pengujian ialah penggunaan metode *naive bayes* saja pada klasifikasi mendapatkan hasil nilai akurasi yang paling optimal sebesar 94,80%. Kemudian sama halnya *Precision* juga mendapatkan hasil nilai sebesar 97,68%., Kemudian *recall* juga mendapatkan hasil nilai sebesar 88,39%., Hal ini dengan demikian dapat meningkatkan hasil nilai akurasi dari algoritma *naive bayes* sehingga menghasilkan akurasi yang lebih optimal. Sistem yang dibangun diharapkan dapat memberi kemudahan serta memperkecil kesalahan yang mungkin terjadi dalam proses pemilihan penerimaan program PKH

Daftar Pustaka

- T. Badan, J. Hb, and M. A. Posteriori, "Klasifikasi Naive Bayes untuk Prediksi Kelahiran pada Data Ibu Hamil," *Bimipa*, vol. 23, no. 3, pp. 297–308, 2013.
- Virgana, dkk., "Kajian Algoritma Naive Bayes dalam Pemilihan Penerimaan Beasiswa Tingkat SMA", Universitas PGRI, 2014
- D. W. Hera wasiati, "Sistem Pendukung Keputusan Penentuan Kelayakan Calon Tenaga Kerja Indonesia Menggunakan Metode Naive Bayes Decision Support System for Determining Eligibility Candidates Indonesian Labor Using Naive Bayes Method (Case Study : Karyatama Mitra Sejati P . T, " 2014.
- P. S. Informatika and F. Teknologi, "Metode Naive Bayes Classier."
- M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 151, 2018, doi: 10.24076/citec.2017v4i2.106.
- D. T. Seimbangan and T. Sampling, "Kombinasi Integrasi Metode Sampling Dengan Naive Bayes Untuk Ketidakseimbangan Kelas Pada," vol. 1, no. 2, pp. 1–7, 2015.
- Fatimah, Gt. Erni, dkk., Penerapan Metode Naive Bayes Dalam Penentuan Bantuan Padat Karya. STMIK Banjarbaru, 2016.
- Nugroho, A., dkk, "Klasifikasi Naive Bayes untuk Prediksi Kelahiran pada Data Ibu Hamil." Berkala MIPA,

2013.

Bustami., (2011), *Penerapan Algoritma Naïve Bayes Untuk Mengklasifikasi Data Nasabah Asuransi*, Universitas Malikussaleh, Aceh.

Kurniawan, A., " *Penentuan Calon Pendorong Darah Menggunakan Algoritma Naïve Bayes Classification*, Universitas Dian Nuswantoro, Semarang.