

# ALGORITMA *NAÏVE BAYES* UNTUK KLASIFIKASI KELAYAKAN PENERIMA BANTUAN IURAN JAMINAN KESEHATAN PEMERINTAH DAERAH KABUPATEN SUKABUMI

Luthfy Budhy Adzy<sup>1</sup>, Asriyanik<sup>2</sup>, Agung Pambudi<sup>3</sup>

<sup>1,2,3</sup> Program Studi Teknik Informatika Fakultas Sains dan Teknologi  
Universitas Muhammadiyah Sukabumi  
luthfybudhy28@ummi.ac.id

## ABSTRAK

Penerima Bantuan Iuran (PBI) Jaminan Kesehatan (JK) merupakan tanggungan berbentuk perawatan kesehatan supaya penerima mendapat utilitas perlindungan kesehatan yang dihibahkan untuk masing-masing masyarakat yang sudah melunasi iuran ataupun iuran itu dibayarkan oleh pemerintah negara. Permasalahan yang biasa terjadi di dalam lapangan yakni dalam pemilihan Keluarga Penerima Manfaat (KPM) yang digunakan masih belum bisa membantu keputusan Supervisor secara objektif dan tepat sasaran. Penelitian ini dilaksanakan untuk membantu Supervisor dalam menentukan cara pemilihan kelayakan calon penerima PBI-JK yang tepat secara objektif dan sesuai target berasaskan standar yang sudah ditetapkan. Pengklasifikasian yang diterapkan mengaplikasikan algoritma *naïve bayes* dengan metode *Knowledge Discovery in Databases* (KDD). Algoritma *naïve bayes* merupakan satu diantara algoritma penggalian data dan pengklasifikasi statistik sebuah klasifikasi berpeluang mudah yang menerapkan teorema bayes melalui asumsi antar variabel ketidak ketergantungan yang luhur. Kelebihan dari algoritma *naïve bayes* adalah bersifat *scalable* dengan jumlah *predictor* dan titik data, bisa membuat prediksi nilai probabilitas (peluang) dan menangani kontinu beserta diskrit data. Pencapaian hasil atas penelitian ini yakni mewujudkan bentuk pengklasifikasian kelayakan penerima bantuan iuran jaminan kesehatan secara otomatis dalam klasifikasi data tersebut layak untuk dibantu atau tidak layak untuk diperbantukan dalam program pemerintah khusus kegiatan Penerima Bantuan Iuran Jaminan Kesehatan Pemerintah Daerah Kabupaten Sukabumi.

**Keyword:** Dinas Sosial, Penerima Bantuan Iuran Jaminan Kesehatan, *Naïve Bayes*, *Knowledge Discovery in Databases*, Keluarga Penerima Manfaat.

## 1. PENDAHULUAN

Perkembangan *data mining* saat ini semakin pesat, membahas mengenai penyaringan data atau penggalian data yang melahirkan satu di antaranya bagian di bidang komputer yang lumayan aktual untuk dipakai dan dialami bagi banyak pakar di bidang pemrogram bahkan menjadi petunjuk pendalaman perihal memecahkan model atau angka. Keberadaan *data mining* dipengaruhi oleh melimpahnya data yang menimpa sebuah instansi. Selain itu *data mining* pun bisa dijabarkan atas proses pemisahan data aktual yang diterima melalui basis data besar (*big data*) untuk mendukung dalam hal mengambil ketetapan atau klasifikasi dengan cara pengelompokan data direalisasikan di sebuah perangkat lunak, yakni proses *semi automatic* dengan menerapkan matematika, statistik, *machine learning*, dan kecerdasan buatan. Salah satu area penerapan ini adalah studi kasus di kantor Dinas Sosial Kabupaten Sukabumi.

Dinas Sosial adalah suatu pelaksana pelaksanaan kegiatan dalam acara pemerintahan dalam aspek kemasyarakatan yang merupakan wewenang wilayah yang memimpin yakni kepala dinas yang memiliki tanggung jawab terhadap Bupati melalui Sekretaris Daerah. Berdasarkan PerMenSos nomor 05 tahun 2016 mengenai pelaksana Peraturan Pemerintah nomor 76 tahun

2015 mengenai transformasi dalam Peraturan Pemerintah nomor 101 tahun 2012 mengenai PBI-JK bahwa yang memiliki wewenang melaksanakan proses mengubah data PBI-JK Pemerintah Daerah (PEMDA) yakni pihak dari dinas sosial kota atau kabupaten. Kemudian dari pihak cabang Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan akan melapor setiap satu bulan sekali kepada pusat BPJS Kesehatan serta surat tembusan ke dinas sosial dan dinas kesehatan. Maka setelah itu, pusat BPJS Kesehatan melapor ke Kementerian Kesehatan.

Selama ini proses klasifikasi kelayakan data penerima bantuan iuran jaminan kesehatan masih dilakukan secara manual oleh Supervisor Dinas Sosial Kabupaten Sukabumi, yakni apabila pihak Desa sudah melakukan musyawarah desa bahwa penerima tersebut layak untuk dibantu maka akan diusulkan ke kantor Dinas Sosial untuk pengusulan PBI-JK PEMDA. Yang dimana program pemerintah khusus tersebut dilakukan satu kali perbulan.

Dari penjelasan permasalahan yang telah dijelaskan di atas, agar proses klasifikasi kelayakan penerima ini sesuai kriteria yang telah ditentukan maka dengan ini peneliti tertarik akan mewujudkan suatu model bentuk pengklasifikasian yang bisa mengakomodasi Supervisor untuk memastikan calon penerima yang tepat sasaran. Dalam klasifikasi kelayakan penerima, hal yang harus

dilakukan untuk menentukan penerima dan bukan penerima terdapat beberapa faktor sebagai bahan pertimbangan dalam penentuan kelayakan penerima di antaranya:

Jurnal yang ditulis oleh Abidin [1] yang menggunakan algoritma *K-Nearest Neighbor* (K-NN) yang berjudul Aplikasi Klasifikasi Penerima Kartu Indonesia Sehat Menggunakan Algoritma *K-Nearest Neighbor*, pada algoritma tersebut mempunyai kekurangan di antaranya memerlukan hitung K selaku indikator, perkiraan jarak *training data* tidak bisa memperjelas dengan bentuk perkiraan yang dipakai, serta estimasi biaya yang cukup tinggi sebab hal tersebut yakni estimasi jarak dari sebagian permintaan data dari basis data untuk seluruh *testing data*.

Jurnal berikutnya yang ditulis oleh Adi [2] yang menggunakan algoritma *naïve bayes classifier* dengan judul Implementasi Algoritma *Naïve Bayes Classifier* untuk Klasifikasi Penerima Beasiswa Peningkatan Prestasi Akademik di Universitas Amikom Yogyakarta, dalam hasil penelitian yang dihasilkan oleh algoritma tersebut yakni mendapatkan hasil nilai *accuracy* 97,66% berikut menggunakan data sampel berjumlah 386 yang tentunya hanya membutuhkan lebih sedikit data *training*. Algoritma *naïve bayes classifier* ini pun mempunyai kelebihan dalam pengklasifikasian di antaranya mudah diimplementasikan, sederhana, cepat, berakurasi tinggi, akan menyatu lebih cepat daripada model diskriminatif seperti regresi logistik, membutuhkan lebih sedikit data *training*, bersifat skalabel atau skala linier dengan jumlah *predictor* serta titik data, bisa membuat prediksi nilai probabilitas dan menangani kontinu beserta diskrit data, dan juga bisa digunakan untuk biner ataupun multi-kelas masalah klasifikasi keduanya.

Berdasarkan perbandingan dua jurnal di atas, dapat peneliti simpulkan bahwa algoritma *naïve bayes* yakni algoritma yang bisa memberikan solusi terhadap permasalahan saat ini dalam proses klasifikasi yang akan dilaksanakan dalam penelitian ini. Atas dasar pertimbangan ini, maka peneliti akan membangun sebuah sistem yang diharapkan dapat membantu penyelesaian masalah mengenai keputusan dalam hal menentukan klasifikasi masyarakat sebagai penerima dan bukan penerima berdasarkan nilai peluang terbesar, tentunya selain itu bagi Supervisor untuk pengelolaan data yang telah masuk setiap satu bulan sekali dari para Puskesmas tiap desa.

## 2. TINJAUAN PUSTAKA

Algoritma Klasifikasi adalah suatu kegiatan memilih atau memberikan label sebuah objek kepada suatu label kategori atau kelompok berdasarkan ciri atau kebutuhan karakter yang telah ditentukan. Dengan pernyataan tersebut, klasifikasi yakni sebuah pemberian label di dalam sebuah objek yang diberikan sebuah kategori tertentu melalui

proses pendefinisian sebuah ciri atau kepemilikan tertentu yang unik sehingga objek itu dapat dikategorikan ke dalam satu keterangan tertentu.

### 2.1. *Naïve Bayes*

Algoritma *naïve bayes* merupakan satu diantara algoritma penggalian data dan pengklasifikasi statistik sebuah klasifikasi berpeluang mudah yang menerapkan teorema bayes dengan melalui asumsi antar variabel-variabel ketidak ketergantungan yang luhur. Implementasi keilmuan umum dan sistem klasifikasi dalam aplikasi sudah banyak diterapkan dalam berbagai kehidupan sehari-hari, misalnya sistem klasifikasi yang digunakan dalam ilmu medis atau kedokteran seperti sistem klasifikasi diagnosis penyakit kronis pada manusia. Atau contoh lain dalam keilmuan umum, seperti pengklasifikasian pakaian berdasarkan warna dalam pemilihan pakaian.

### 2.2. Langkah-langkah Algoritma

Bentuk persamaan atau umum dari teorema bayesian, berikut ini:

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (1)$$

Berikut terdapat beberapa keterangan pada rumus (1) di atas, di antaranya X artinya data dengan kelas yang belum diketahui, C artinya hipotesis data X adalah sebuah kelas spesifik,  $P(C|X)$  artinya probabilitas hipotesis C berdasarkan kondisi X,  $P(C)$  artinya probabilitas hipotesis C (probabilitas prior),  $P(X|C)$  artinya probabilitas X berdasarkan kondisi pada hipotesis C, dan  $P(X)$  artinya probabilitas X.

Pengimplementasian menggunakan metode *naïve bayes* bertujuan untuk mengambil angka peluang atas dasar variabel-variabel label bagi tiap variabel ketidak ketergantungan yang kemudian bisa ditunjukkan layak sebagai penerima, menggunakan data atas perolehan instansi tersebut. Maka dari itu, dilaksanakan pencarian rentang nilai secara perulangan dari *training data* yang diterima oleh peneliti.

Setelah tahap ini selesai, maka tahap selanjutnya pengujian nilai akurasi, *precision*, serta *recall* terhadap keluaran yang akan dihasilkan oleh aplikasi. Akurasi merupakan kalkulasi dengan mendapatkan keluaran dari jumlah perkiraan, berikut rumus untuk perhitungan nilai *accuracy* di antaranya:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

Kemudian, *precision* merupakan kalkulasi untuk mencari keluaran diagnosis positif, berikut rumus untuk perhitungan nilai *precision* di antaranya:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

Lalu, *recall* merupakan kalkulasi untuk mencari keluaran diagnosis positif dengan teridentifikasi benar, berikut rumus untuk perhitungan nilai *recall* di antaranya:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Yang terakhir, *f1-score* merupakan penggabungan *precision* dan *recall* menjadi matriks tunggal dengan mengambil rata-rata harmonis, berikut rumus untuk perhitungan nilai *f1-score* di antaranya:

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \quad (5)$$

Berikut keterangan dari rumus (2), (3), dan (4) di atas, di antaranya TP artinya *True Positive*, lalu TN artinya *True Negative*, kemudian FP artinya *False Positive*, dan yang terakhir FN artinya *False Negative*. Tahap terakhir yakni evaluasi tingkat *accuracy* dari metode tersebut, dengan *tools* yang dipakai yaitu perangkat lunak RapidMiner dengan menggunakan model *confusion matrix* dan kurva *Receiver Operating Characteristic (ROC)*.

**2.3. Rapidminer**

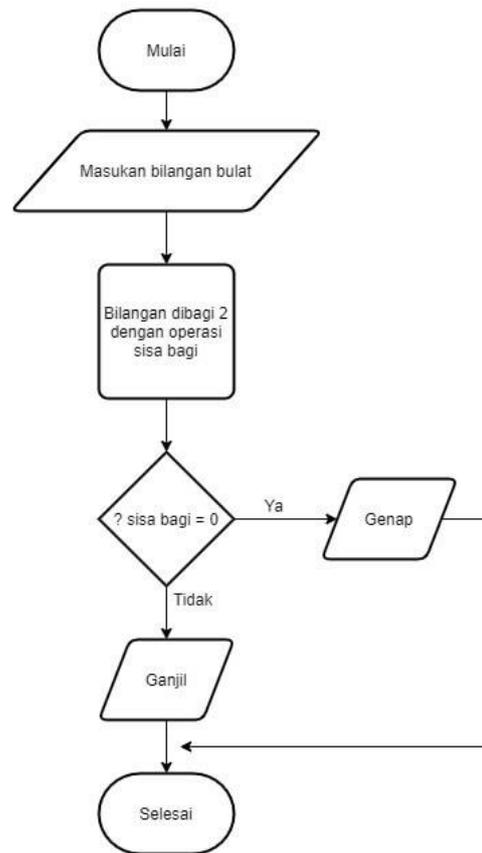
Rapidminer adalah perangkat lunak (*software*) atau *data mining tools* yang cocok dalam analisis prediktif terbaik yang dapat digunakan untuk riset penelitian, pengembangan aplikasi, dan lain sebagainya [3].

**2.4. Visual Studio Code (VSCode)**

Kode editor VSCode atau kependekan dari *Visual Studio Code* adalah sebuah perangkat lunak untuk mengedit kode pemrograman yang didukung oleh berbagai jenis sistem operasi dan berbagai macam bahasa pemrograman [4].

**2.5. Flowchart**

*Flowchart* merupakan bentuk bagan dengan berbagai macam pola dengan simbol-simbol yang mempunyai jenis dan fungsinya masing-masing serta arti yang berbeda-beda sehingga hal tersebut memberikan suatu gambaran bagaimana berjalannya sebuah sistem yang akan dibangun oleh peneliti nantinya, maka alur sistem tersebut menjadi mudah untuk dipahami oleh siapapun yakni dengan menyederhanakan rangkaian prosedur penelitian yang dibuat oleh peneliti [5]. Berikut contoh dari sebuah *flowchart* sederhana, di antaranya:



Gambar 1. Contoh *Flowchart* [5]

Berikut terdapat beberapa simbol keterangan pada gambar 1 di atas, di antaranya simbol *flowline* berfungsi sebagai penghubung antar simbol dengan yang lainnya, simbol *process* berfungsi sebagai pernyataan sebuah proses yang dilakukan oleh sistem, simbol *terminal* yang berfungsi sebagai pernyataan sebuah awalan atau akhiran dari program, simbol *input / output* berfungsi sebagai pernyataan sebuah proses masukan dan keluaran tanpa tergantung pada peralatan, kemudian yang terakhir simbol *decision* berfungsi sebagai pernyataan kondisi tertentu yang nantinya akan menghasilkan dua buah kemungkinan jawaban antara ya dan tidak.

**2.6. Hypertext Preprocessor (PHP)**

Bahasa pemrograman PHP atau kependekan dari *Hypertext Preprocessor* yakni pemrograman bahasa komputer yang dipakai dalam hal membangun sebuah *website* yang berjalan dengan menyisipkan Javascript, CSS, Bootstrap, dan lain sebagainya [6].

**2.7. Website**

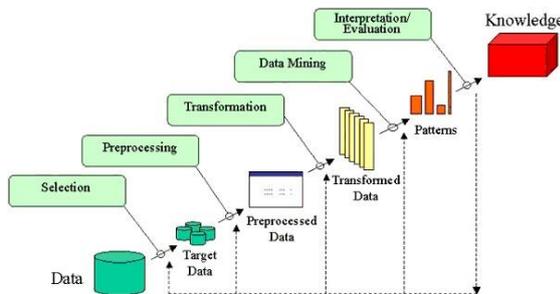
*Website* merupakan sekumpulan halaman *web* yang saling berhubungan pada peladen yang sama berisikan informasi yang dapat diakses oleh banyak orang secara per-individu, berkelompok, ataupun organisasi melalui koneksi internet [7].

**2.8. Basis Data dan My Structured Query Language (MySQL)**

Basis data dan MySQL atau kependekan dari merupakan suatu metode untuk mengatur tata laksana *database* relasional dengan membuat relasi antar tabel [8], MySQL bisa dikatakan sebuah tempat khusus untuk menyimpan data secara terpisah dan tersusun dalam pengelolaan data tersebut dengan menggunakan aplikasi atau program untuk menghasilkan suatu sistem [9].

**2.9. Metode Knowledge Discovery in Databases (KDD)**

Metode KDD atau kependekan dari *Knowledge Discovery in Databases* merupakan keutuhan teknik dalam hal menggali serta mengenal suatu pola dari sekumpulan data yang dimana pola tersebut berkarakter modern, dan juga bisa berguna serta bisa dipahami. Berikut gambar tahapan-tahapan dari metode KDD di antaranya:



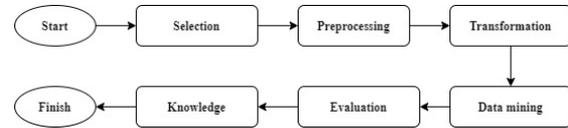
Gambar 2. Metode KDD

Berikut terdapat beberapa keterangan pada gambar 2 di atas, di antaranya data *selection* yakni proses seleksi atau pemilihan data yang dianggap relevan terhadap analisis [10], target data / *preprocessing* yakni proses dimana data diolah lalu dipilih data yang dianggap bisa dipakai [10], *preprocessed data* / *transformation* yakni proses transformasi data terpilih ke dalam bentuk *mining procedure* [10], *transformed data* / *data mining* yakni proses dimana dilakukan berbagai teknik untuk mengekstrak pola-pola potensial menghasilkan data yang berguna [10], *pattern* / *evaluation* yakni proses dimana pola-pola yang telah diidentifikasi berdasarkan *measure* yang diberikan [10], kemudian yang terakhir *knowledge* yakni proses paling akhir dari proses KDD serta informasi yang sudah diproses divisualisasikan agar lebih mudah dipahami oleh pengguna dan diharapkan bisa diambil tindakan berdasarkan analisis [10].

**3. METODE PENELITIAN**

Pelaksanaan dari penelitian ini diajukan untuk membantu penyelesaian masalah mengenai keputusan dalam menentukan klasifikasi masyarakat penerima dan bukan penerima berdasarkan nilai peluang terbesar. Dengan metode KDD menggunakan algoritma *naive bayes* yakni salah satu metode pemecahan masalah dalam

pengolahan data tersebut, tahapan penelitian ini dimulai dari *selection*, *preprocessing*, *transformation*, *data mining*, *evaluation*, dan *knowledge*. Di antaranya tahapan yang akan dilaksanakan atas penelitian tersebut, bisa diketahui dalam gambar di bawah ini:



Gambar 3. Tahapan Penelitian

Berikut penjelasan dari gambar 3 tahapan penelitian di atas yang digunakan oleh peneliti, di antaranya:

**3.1. Selection**

Proses seleksi atau pemilihan data merupakan langkah awal dalam pemrosesan perhitungan klasifikasi dengan algoritma *naive bayes* dalam penelitian ini yakni mengumpulkan data-data apa saja yang dibutuhkan untuk dilakukan proses klasifikasi kelayakan, salah satunya adalah peneliti melakukan wawancara kepada seorang supervisor dinas sosial kabupaten sukabumi dan data tambahan yang berasal dari wawancara bersama salah seorang puskesmas dinas sosial yang ada di desa tersebut. Berikut data perolehan dari hasil wawancara, di antaranya:

Tabel 1. Data Asli

No.	Data yang dipakai	Narasumber	Teknik Pengumpulan Data
1.	Data KPM	Supervisor Dinas Sosial	Wawancara

Pada tabel 1 di atas, membuktikan bahwa data primer yang digunakan oleh peneliti yakni data keluarga penerima manfaat dari hasil wawancara dengan pihak supervisor dinas sosial kabupaten sukabumi. Didalam tahap teknik pengumpulan data ini, yang dimana peneliti bertanya dengan beberapa point-point penting di antaranya salah satunya menanyakan seputar data-data masyarakat dari masing-masing tiap desa yang terdiri dari 386 desa di wilayah kabupaten sukabumi.

**3.2. Preprocessing**

Data pra-pemrosesan atau pemrosesan awal merupakan tahap *cleaning data* dengan memeriksa duplikat data serta tingkat konsistensi, yang dimana tujuan dari proses ini yakni mengubah data mentah menjadi informasi yang sebanding atas apa yang akan dianalisis selanjutnya oleh peneliti. *Preprocessing* data ini tentunya bisa menaikkan kualitas data dan nilai *accuracy* serta kemampuan teknik penggalan data.

**3.3. Transformation**

Data transformasi atau perubahan data yakni dengan melakukan perubahan tipe data menjadi sesuai ketentuan, sehingga data siap untuk disajikan ke dalam proses penggalian data. Jadi, data transformasi tersebut bertujuan untuk mengubah tipe data menjadi bentuk standar dengan apa yang akan dilakukan dalam penambangan data nantinya.

**3.4. Data Mining**

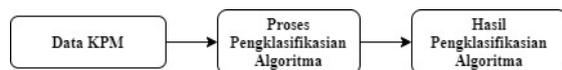
Penggalian data atau penyaringan data yakni sebuah proses pemilihan tugas sebanding dengan tujuan dan algoritma *naive bayes* ini untuk mencari sebuah pola pemodelan data. Dari data *preparation* yang telah didapatkan oleh peneliti, selanjutnya akan dilakukan pemodelan algoritma tersebut dengan model peluang *naive bayes* dalam nilai peluang itu dilakukan perhitungan dengan dimasukkan ke dalam *tools* RapidMiner.

**3.5. Evaluation**

Proses evaluasi atau evaluasi data merupakan langkah terakhir dalam pemrosesan perhitungan klasifikasi dengan algoritma itu dalam penelitian ini yakni pengujian terhadap nilai *accuracy*. Maka proses pengujiannya dilakukan dengan menggunakan *tools* yang dipakai adalah perangkat lunak RapidMiner.

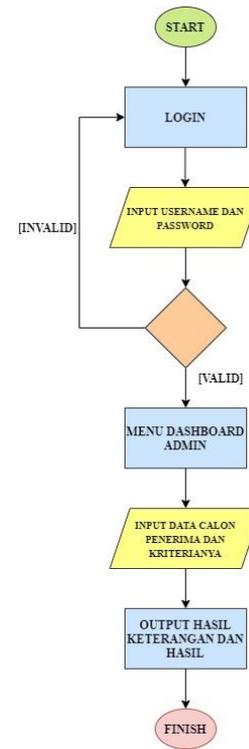
**3.6. Knowledge**

Presentasi pengetahuan atau sistem aplikasi klasifikasi penerima bantuan iuran jaminan kesehatan dibangun memakai bahasa pemrograman PHP dan basis data MySQL serta *tools* RapidMiner, hasil dari implementasi sistem ini merupakan sistem berbasis *web* yang bisa dipakai untuk proses mengklasifikasikan data KPM. Berikut tahapan implementasi pada sistem aplikasi klasifikasi PBI-JK, di antaranya:



Gambar 4. Tahapan Implementasi pada Sistem Aplikasi PBI-JK PEMDA

Berikut terdapat beberapa keterangan pada gambar 4 di atas, di antaranya berawal dari *data set* yang dikumpulkan atau dipakai dalam pelaksanaan penelitian ini yang kemudian dimasukkan ke dalam proses pengklasifikasian algoritma. Kemudian langkah terakhir yakni keluarnya hasil dari proses pengklasifikasian algoritma tersebut. Selain itu juga terdapat alur *flowchart* dari sistem aplikasi ini, di antaranya:



Gambar 5. Flowchart Sistem Aplikasi PBI-JK PEMDA

Berikut terdapat beberapa simbol keterangan pada gambar 5 di atas, di antaranya simbol *flowline* berfungsi sebagai penghubung antar simbol dengan yang lainnya, simbol *process* berfungsi sebagai pernyataan sebuah proses yang dilakukan oleh sistem, simbol *terminal* yang berfungsi sebagai pernyataan sebuah awalan atau akhiran dari program, simbol *input / output* berfungsi sebagai pernyataan sebuah proses masukan dan keluaran tanpa tergantung pada peralatan, kemudian yang terakhir simbol *decision* berfungsi sebagai pernyataan kondisi tertentu yang nantinya akan menghasilkan dua buah kemungkinan jawaban antara ya dan tidak atau *valid* dan *invalid*.

**4. HASIL DAN PEMBAHASAN**

**4.1. Pengumpulan Data Set**

Peneliti mengumpulkan sebuah informasi dengan melakukan wawancara bersama supervisor dinas sosial kabupaten sukabumi. *Data set* yang dipakai dalam pelaksanaan penelitian ini terdiri atas data masyarakat PBI-JK PEMDA Kabupaten Sukabumi TA. 2021.

Jumlah *data set* asli pada data masyarakat terdiri dari 1.132.715 jiwa / *dataset*, dari keseluruhan atribut di bawah yang akan digunakan dalam proses KDD hanya 12 atribut. Berikut atribut-atribut dalam *data set* tersebut dan jumlah *data set* di dalam *tools* RapidMiner, di antaranya:

<b>ID DTKS</b>	Menunjukkan kode atau ID masyarakat
<b>Provinsi</b>	Menunjukkan dimana letak provinsi
<b>Kabupaten</b>	Menunjukkan dimana letak kabupaten
<b>Kecamatan</b>	Menunjukkan dimana letak kecamatan
<b>Desa</b>	Menunjukkan dimana letak desa
<b>RT</b>	Menunjukkan alamat RT
<b>RW</b>	Menunjukkan alamat RW
<b>Luas lantai</b>	Menunjukkan ukuran luas lantai bangunan
<b>Jenis lantai</b>	Menunjukkan jenis lantai tempat tinggal
<b>Jenis dinding</b>	Menunjukkan jenis dinding tempat tinggal
<b>Fasilitas buang air besar</b>	Menunjukkan kepemilikan fasilitas buang air besar
<b>Sumber penerangan</b>	Menunjukkan sumber penerangan tempat tinggal
<b>Sumber air</b>	Menunjukkan sumber air berasal
<b>Bahan bakar memasak</b>	Menunjukkan bahan bakar untuk memasak
<b>Porsi makan per hari</b>	Menunjukkan kesanggupan porsi makan per hari
<b>Pendapatan</b>	Menunjukkan sumber penghasilan kepala keluarga
<b>Pendidikan terakhir</b>	Menunjukkan pendidikan terakhir kepala keluarga
<b>Status</b>	Menunjukkan status kelayakan sebagai penerima

Gambar 6. Data Attribute PBI-JK PEMDA TA. 2021

Berikut terdapat beberapa keterangan pada gambar 6 di atas, di antaranya baris yang berwarna kuning mengartikan sebuah label atribut dan label target pada *data set* tersebut serta menunjukkan fungsinya masing-masing, lalu baris yang berwarna biru mengartikan bahwa atribut data yang akan dilakukan *data selection / preprocessing* data dalam proses KDD nantinya serta menunjukkan fungsinya masing-masing, kemudian yang terakhir baris berwarna hijau mengartikan sebuah atribut pendukung pada *data set* tersebut serta menunjukkan fungsinya masing-masing.

Gambar 7. Data Set PBI-JK PEMDA TA. 2021

Berikut terdapat keterangan pada gambar 7 di atas, menunjukkan bahwa seluruh *data set* PBI-JK PEMDA TA. 2021 berjumlah 1.048.574 *data set* telah dimasukkan ke dalam *tools* RapidMiner yang kemudian akan dilanjutkan pada proses *preprocessing* data nantinya.

**4.2. Preprocessing**

Dalam tahapan ini, menunjukkan hasil *preprocessing* data. Yang dimana hasil penyeleksian / *selection* data mengalami *preprocessing* yakni dilakukan data yang tidak relevan, tidak konsisten, *missing value*, pembersihan data-data redundansi / data duplikat, ataupun menghapus data yang tidak digunakan. Berikut hasil seleksi data yang dihasilkan oleh *tools* RapidMiner di antaranya:

Gambar 8. Selection Data PBI-JK PEMDA TA. 2021

Berikut terdapat keterangan pada gambar 8 di atas, menunjukkan bahwa hal tersebut hasil dari proses *selection* data PBI-JK PEMDA TA. 2021 yang dimana dengan mengecek dan menghapus atribut data yang tidak digunakan di antaranya “Provinsi”, “Kabupaten”, “Kecamatan”, “Desa”, “RT”, dan “RW”. Karena data atau atribut-atribut tersebut tidak dijadikan sebagai acuan dalam proses klasifikasi ini.

**4.3. Transformation**

Data transformasi atau perubahan data merupakan teknik mengubah data dalam beberapa jenis atribut yang sudah ditetapkan menjadi bentuk yang diinginkan dalam (*input – process – output*), sehingga jenis atribut itu sudah sesuai dengan teknik pengalihan data untuk kebutuhan analisa nantinya. Pada data masyarakat ini, apabila sudah dilakukan proses *preprocessing* atau *cleaning* maka selanjutnya ditransformasikan. Pada tahapan *transformation* dalam KDD yakni proses kreativitas dan sangat bergantung terhadap *pattern* atau pola informasi yang dihasilkan oleh *tools* RapidMiner di antaranya:

Tabel 2. Data Transformation atribut Luas Lantai

Luas Lantai	Transformation
< 8 m <sup>2</sup>	sempit
> 8 m <sup>2</sup>	luas

Tabel 3. Data Transformation atribut Jenis Lantai

Jenis Lantai	Transformation
tanah	murah
bambu murah	
kayu murah	
keramik	mahal
plester	

Tabel 4. Data Transformation atribut Jenis Dinding

Jenis Dinding	Transformation
bambu murah	murah
rumbia	
kayu murah	
tembok tanpa dipleser	
GRC	mahal
tembok	

Tabel 5. Data Transformation atribut Sumber Air

Sumber Air	Transformation
sumur tradisional	gratis
mata air	
sungai	
air hujan	
PDAM	berbayar

Tabel 6. Data Transformation atribut Bahan Bakar Memasak (BBM)

BBM	Transformation
kayu bakar	murah
arang	
minyak tanah	
LPG	mahal

Tabel 7. Data Transformation atribut Porsi Makan

Porsi Makan	Transformation
1 – 2 kali sehari	sedikit
> 2 kali sehari	banyak

Tabel 8. Data Transformation atribut Pendapatan

Pendapatan	Transformation
< 600.000 per bulan	kecil
> 600.000 per bulan	besar

Tabel 9. Data Transformation atribut Pendidikan Terakhir

Pendidikan Terakhir	Transformation
tidak sekolah	rendah
SD	
SLTP	
SLTA	tinggi

Berikut terdapat keterangan dari beberapa tabel 2, 3, 4, 5, 6, 7, 8, dan 9 di atas adalah sebuah hasil dari proses transformation data yang dimana untuk meningkatkan keandalan data dengan teknik melakukan perubahan isi dataset ke dalam bentuk kategori, sehingga hal tersebut siap untuk dilanjutkan pada proses data mining nantinya.

No	Kecamatan	Desa	Jenis Lahan	Jenis Dringku	Fasilitas BUMDes	Sumber Pangan	Sumber Air	Bahan Bakar Memasak	Porsi Makan
1	PRINGSELO	SEMPIT	MUSIAH	MUSIAH	UMUR	NON-LESTARI	UMUR	SEDIKIT	SEDIKIT
2	PRINGSELO	LUAR	MUSIAH	MUSIAH	PRINGSELO	NON-LESTARI	UMUR	MANYAK	MANYAK
3	PRINGSELO	LUAR	MUSIAH	MUSIAH	PRINGSELO	NON-LESTARI	UMUR	MANYAK	MANYAK
4	PRINGSELO	LUAR	MUSIAH	MUSIAH	UMUR	LESTARI	UMUR	MANYAK	MANYAK
5	PRINGSELO	LUAR	MUSIAH	MUSIAH	UMUR	LESTARI	UMUR	SEDIKIT	SEDIKIT
6	PRINGSELO	LUAR	MUSIAH	MUSIAH	UMUR	NON-LESTARI	UMUR	MANYAK	SEDIKIT
7	PRINGSELO	LUAR	MUSIAH	MUSIAH	UMUR	NON-LESTARI	UMUR	MANYAK	SEDIKIT
8	PRINGSELO	LUAR	MUSIAH	MUSIAH	UMUR	NON-LESTARI	UMUR	MANYAK	SEDIKIT
9	PRINGSELO	LUAR	MUSIAH	MUSIAH	UMUR	NON-LESTARI	UMUR	MANYAK	MANYAK
10	PRINGSELO	LUAR	MUSIAH	MUSIAH	PRINGSELO	LESTARI	UMUR	MANYAK	SEDIKIT
11	PRINGSELO	SEMPIT	MUSIAH	MUSIAH	UMUR	NON-LESTARI	UMUR	MANYAK	MANYAK
12	PRINGSELO	LUAR	MUSIAH	MUSIAH	UMUR	NON-LESTARI	UMUR	MANYAK	SEDIKIT
13	PRINGSELO	LUAR	MUSIAH	MUSIAH	UMUR	NON-LESTARI	UMUR	MANYAK	SEDIKIT
14	PRINGSELO	LUAR	MUSIAH	MUSIAH	UMUR	NON-LESTARI	UMUR	MANYAK	SEDIKIT
15	PRINGSELO	LUAR	MUSIAH	MUSIAH	UMUR	NON-LESTARI	UMUR	MANYAK	SEDIKIT

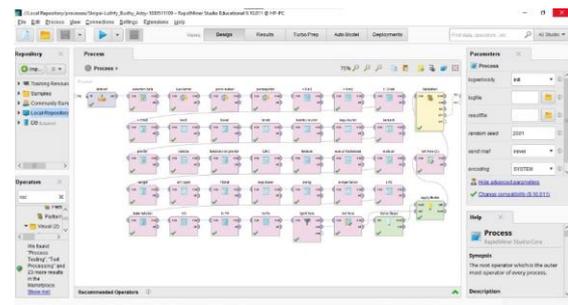
Gambar 9. Transformation Data PBI-JK PEMDA TA. 2021

Berikut terdapat keterangan pada gambar 9 di atas, menunjukkan bahwa hal tersebut hasil dari

proses transformation data PBI-JK PEMDA TA. 2021 berdasarkan keterangan dari beberapa tabel di atas. Proses ini dilakukan dalam tools RapidMiner yang kemudian akan dilanjutkan pada proses data mining nantinya.

4.4. Pemodelan Data

RapidMiner Studio 9.10. adalah salah satu tools data mining untuk melakukan sebuah prediksi, klasifikasi, dan analisis data, berikut ini pengujian pemodelan data yang dilakukan pada 1.048.574 data set (80% training data dan 20% testing data) di antaranya:



Gambar 10. Process Pemodelan Data PBI-JK PEMDA TA. 2021

Berikut terdapat beberapa keterangan pada gambar 10 di atas, di antaranya operators “Read excel” untuk memasukkan data set itu, operators “Select attributes” untuk menentukan atribut yang akan digunakan dalam proses klasifikasi, operators “Numerical to Binomial” untuk mengubah data numerik menjadi data kategorial, operators “Replace” untuk mengganti isi data sesuai pengubahan data yang telah ditentukan, operators “Split Data” untuk mengambil jumlah data training dan data testing yang akan diolah dalam proses klasifikasi, operators “Set Role” untuk menentukan atribut pendukung dari data set itu, operators “Naive Bayes” untuk proses klasifikasi menggunakan algoritma itu, operators “Apply Model” untuk menerapkan model data yang telah dilatih sebelumnya, operators “Cross Validation” untuk melatih model dan sub-proses testing untuk pengujian serta mengukur kinerja pemodelan data itu.

4.5. Hasil Process Pemodelan Data

Berdasarkan proses pemodelan data menggunakan tools RapidMiner ini, berikut hasil pemrosesan menggunakan algoritma naive bayes dengan mendapatkan nilai accuracy sebesar 96,83% di antaranya:

Class	Actual	Predicted	Count	Percentage
layak	layak	layak	114300	87.0%
	layak	tidak layak	3166	2.4%
tidak layak	tidak layak	tidak layak	88763	68.2%
	tidak layak	layak	3486	2.7%

Gambar 11. Nilai *Performance Vector* Data PBI-JK PEMDA TA. 2021

Berikut terdapat beberapa keterangan pada gambar 11 di atas, di antaranya *True Positive* (TP) sebanyak 114.300 record yang diklasifikasikan sebagai penerima “layak”, *False Negative* (FN) sebanyak 3.166 record yang diklasifikasikan sebagai penerima “tidak layak”, *True Negative* (TN) sebanyak 88.763 record yang diklasifikasikan sebagai penerima “tidak layak”, *False Positive* (FP) sebanyak 3.486 record yang diklasifikasikan sebagai penerima “layak”, nilai *precision* sebesar 97,30% pada prediksi “layak” dan 96,22% pada prediksi “tidak layak”, nilai *recall* sebesar 97,04% pada prediksi “layak” dan 96,56% pada prediksi “tidak layak”, nilai *f1-score* sebesar 97% pada penggabungan nilai *precision* dan nilai *recall*.

**4.6. Evaluation**

Berikut kurva ROC yang digunakan untuk mengekspresikan data, di antaranya:



Gambar 12. Nilai AUC dalam Grafik ROC Data PBI-JK PEMDA TA. 2021

Berikut terdapat keterangan pada gambar 12 di atas menunjukkan bahwa dengan menggunakan kurva ROC untuk mengekspresikan data tersebut, dengan perolehan nilai *Area Under Curve* (AUC) yakni senilai 0,988 ini menunjukkan bahwa pemodelan data tersebut mencapai klasifikasi hampir sempurna.

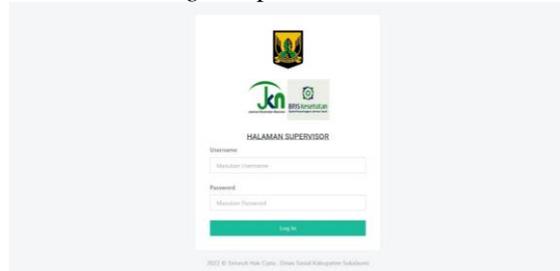
**4.7. Knowledge**

Langkah terakhir dalam pelaksanaan penelitian ini yakni pembentukan *knowledge* dari hasil pemodelan data pada aplikasi *web* PBI-JK PEMDA menggunakan algoritma *naïve bayes* dengan *tools* RapidMiner, adapun hasilnya seperti yang ditunjukkan pada gambar 10 di atas.

**4.8. Implementasi Antarmuka**

Berikut hasil implementasi antarmuka aplikasi yang dibuat oleh peneliti, di antaranya:

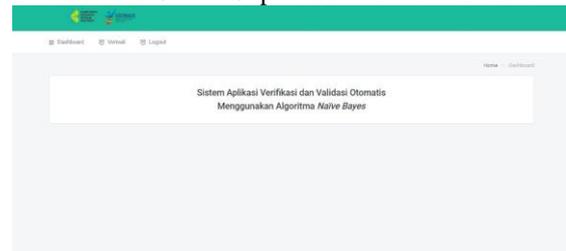
1. Halaman *Login* Supervisor PBI-JK PEMDA



Gambar 13. Tampilan Halaman *Login* Supervisor PBI-JK PEMDA

Berikut terdapat keterangan pada gambar 13 di atas, menunjukkan bahwa halaman *login* tersebut diperuntukkan bagi supervisor dinas sosial kabupaten sukabumi. Pada proses ini dipakai untuk sebelum masuk ke halaman utama supervisor nantinya, dengan memasukkan *username* dan *password* yang *valid*.

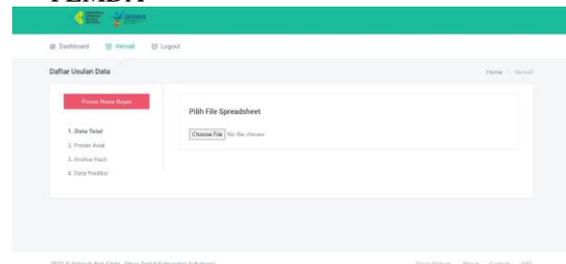
2. Halaman Utama Supervisor PBI-JK PEMDA



Gambar 14. Tampilan Halaman Utama Supervisor PBI-JK PEMDA

Berikut terdapat keterangan pada gambar 14 di atas, menunjukkan bahwa halaman utama admin diperuntukkan bagi supervisor dinas sosial kabupaten sukabumi. Pada proses ini setelah *login* berhasil, maka akan masuk ke halaman tersebut yang berisikan menu *dashboard*, *verivali*, dan *logout*.

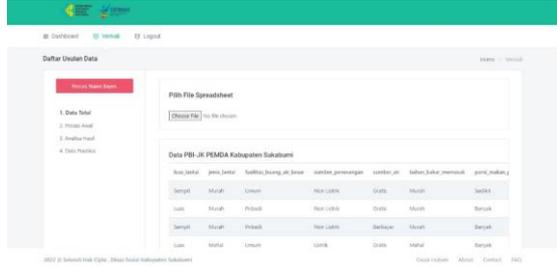
3. Halaman Verifikasi dan Validasi PBI-JK PEMDA



Gambar 15. Tampilan Halaman Verifikasi dan Validasi PBI-JK PEMDA

Berikut terdapat keterangan pada gambar 15 di atas, menunjukkan bahwa halaman verifikasi dan validasi data. Pada proses ini dilakukan dengan cara *import file spreadsheet* berupa tabel dengan format yang telah ditentukan.

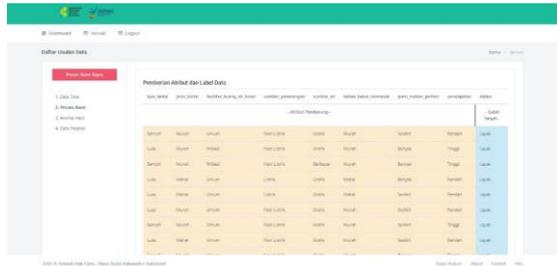
4. Halaman Total Data PBI-JK PEMDA



Gambar 16. Tampilan Halaman Total Data PBI-JK PEMDA

Berikut terdapat keterangan pada gambar 16 di atas, menunjukkan bahwa halaman total data. Pada proses ini setelah *import file spreadsheet* atau memasukkan *data set* tersebut, maka isi keseluruhan dari *file* itu akan muncul pada halaman aplikasi *web* tersebut.

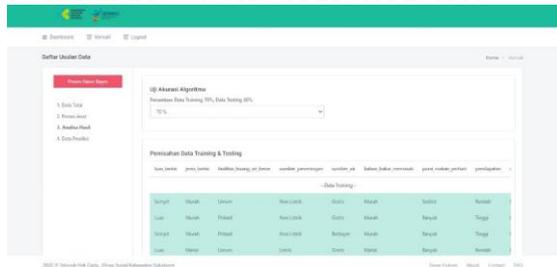
5. Halaman Proses Awal PBI-JK PEMDA



Gambar 17. Tampilan Halaman Proses Awal PBI-JK PEMDA

Berikut terdapat keterangan pada gambar 17 di atas, menunjukkan bahwa proses awal. Pada proses ini akan muncul setelah *import file spreadsheet* atau memasukkan *data set* tersebut berhasil, maka akan otomatis dilakukan pemberian atribut dan label data disebut dengan atribut pendukung dan label target.

6. Halaman Analisa Hasil PBI-JK PEMDA

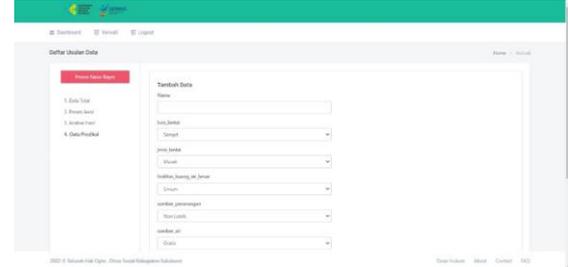


Gambar 18. Tampilan Halaman Analisa Hasil PBI-JK PEMDA

Berikut terdapat keterangan pada gambar 18 di atas, menunjukkan bahwa analisa hasil. Pada proses

ini dilakukan dengan cara memilih persentase data *training* dan data *testing* pada menu uji akurasi algoritma tersebut, maka setelah itu akan muncul nilai *accuracy*, *precision*, dan *recall* dari hasil proses sistem pengklasifikasian tersebut.

7. Halaman Prediksi Data PBI-JK PEMDA



Gambar 19. Tampilan Halaman Prediksi Data PBI-JK PEMDA

Berikut terdapat keterangan pada gambar 19 di atas, menunjukkan halaman prediksi data. Pada proses ini dilakukan dengan cara mengisi form-form yang telah disediakan pada menu tambah data, setelah itu maka akan muncul hasil perhitungan dengan menggunakan rumus teorema bayes untuk menentukan layak dan tidak data masyarakat tersebut untuk menerima bantuan PBI-JK PEMDA Kabupaten Sukabumi.

4.9. Implementasi Algoritma

Pada implementasi algoritma ini adalah penerapan algoritma *naive bayes* ke dalam sebuah koding atau sistem, berikut di antaranya:

```
<?php
defined('BASEPATH') OR exit('No direct script access allowed');
include ("libraries/autoload.php");
use GroceryCrud\Core\GroceryCrud;
class NaiveBayes extends CI_Controller{
var $footer = [];
var $menu = [];
function __construct(){
parent::__construct();
$dbdatabase = include ('database.php');
$config = include ('config.php');
$this->crud = new GroceryCrud($config, $dbdatabase);
$this->crud->unsetBootstrap();
$this->crud->unsetExport();
$this->crud->unsetPrint();
$this->footer = array(
"Dasar Hukum"=>"#LinktoDasarHukum",
"About"=>"#LinktoAbout",
"Contact"=>"#LinktoContact",
"FAQ"=>"#LinktoFAQ"
);
$this->menu = array(
"navbas"=>array(
"menu"=>array(
array(
"name"=>"Dashboard",
"icon"=>"remixicon-dashboard-line",
"link"=>base_url()."naivebayes"
),
array(
"name"=>"Verivali",
"icon"=>"remixicon-honour-line",
"link"=>base_url()."naivebayes/process"
),
array(
"name"=>"Logout",
"icon"=>"remixicon-honour-line",
"link"=>base_url()."auth/logout"
),
)
)
);
};
};
```

Kode Program 1. Controller NaiveBayes

Berikut terdapat keterangan pada kode program 1 di atas, menunjukkan controller *naive*

bayes. Pada *script* koding ini peneliti mengimplementasikan algoritma *naive bayes* ke dalam bahasa pemrograman PHP dengan menggunakan *framework CodeIgniter 4* yakni untuk proses pengklasifikasian data tersebut yang berisikan *class NaiveBayes*.

**4.10. Pengujian Sistem**

Pada pengujian sistem ini dilaksanakan untuk menguji apakah sistem tersebut layak digunakan atau kurang layak. Maka dari itu, dilakukanlah sebuah pengujian sistem dengan menggunakan metode *black box testing*. Berikut ini gambar tabel pengujian sistem, di antaranya:

 Universitas Muhammadiyah Sukabumi Fakultas Sains dan Teknologi Program Studi Teknik Informatika					
<b>LEMBAR PENGUJIAN ALGORITMA</b>					
Nama	: Luthfy Budhy Adzy				
Nomer Induk Mahasiswa	: 1830511109				
Program Studi	: Teknik Informatika				
Jenjang Studi	: Sarjana				
Judul Skripsi	: Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Iuran Jaminan Kesehatan Pemerintah Daerah Kabupaten Sukabumi				
Nama Penguji	: Pepi S Wiguna, S.T.				
Jabatan	: Operator SIKS-NG PBI				
<b>Rencana dan Hasil Pengujian</b>					
No.	Modul/Bagian yang diuji	Alur Pengujian	Hasil yang diharapkan	Kondisi yang terjadi pada saat pengujian	Kesimpulan
1.	Login Supervisor Dinas Sosial	✓	✓	✓	SUKSES
2.	Import Data PBI-JK	✓	✓	✓	SUKSES
3.	Penambahan Data Total	✓	✓	✓	SUKSES
4.	Penambahan Data Training	✓	✓	✓	SUKSES
5.	Penambahan Data Testing	✓	✓	✓	SUKSES
6.	Penambahan Data Prediksi	✓	✓	✓	SUKSES
7.	Memampilkan Hasil Data Analisa	✓	✓	✓	SUKSES
8.	Logout Supervisor Dinas Sosial	✓	✓	✓	SUKSES

Gambar 20. Lembar Pengujian Sistem

Berikut terdapat keterangan pada gambar 20 di atas, menunjukkan bahwa pengujian sistem ini berharap aplikasi ini dapat menyelesaikan permasalahan yang sedang diteliti oleh peneliti dan menjamin kelayakan, kenyamanan, serta dapat memenuhi kebutuhan aplikasi PBI-JK PEMDA berbasis *web* ini. Pada pengujian sistem tersebut peneliti melakukan pengujian secara langsung di dalam aplikasi tersebut bersama dengan pihak instansi terkait, yang dimana nantinya aplikasi tersebut untuk dapat digunakan oleh Supervisor Dinas Sosial Kabupaten Sukabumi.

**5. KESIMPULAN DAN SARAN**

Dari pembahasan atau penjelasan yang sudah dilaksanakan didalam penelitian tersebut, bisa diterima kesimpulan bahwa hasil pemodelan data pada penelitian yang dilaksanakan. Dari pengujian data yang dilakukan dengan membandingkan data *training* dan *testing* menggunakan *tools RapidMiner* didapatkan tingkat akurasi sebesar 96,83%.

Selain itu juga, terdapat saran untuk mengembangkan penelitian selanjutnya yakni

dengan memperbaiki sistem aplikasi dengan cara mencari bahan acuan program sehingga dapat lebih mudah dalam proses pengerjaan dan sistem klasifikasi otomatis yang lebih sempurna. Yang dimana hasil pemodelan data tersebut tanpa harus melalui *tools RapidMiner* terlebih dahulu.

**DAFTAR PUSTAKA**

- [1] A. Z. Abidin, “Aplikasi Klasifikasi Penerima Kartu Indonesia Sehat Menggunakan Algoritma K-Nearest Neighbor,” *publikasi.dinus.ac.id*, vol. 4, pp. 151–161, 2019, doi: 10.26905/jtmi.v4i1.1870.
- [2] S. Adi, “Implementasi Algoritma Naive Bayes Classifier Untuk Klasifikasi Penerima Beasiswa PPA Di Universitas Amikom Yogyakarta,” *jurnal.ubl.ac.id*, vol. 22, no. 1, pp. 11–16, 2018, [Online]. Available: <http://e-jurnal.pelitanusantara.ac.id/index.php/mantik/article/view/342>.
- [3] A. Sotech, “Pilihan Data Mining Tools yang Penting untuk Perusahaan Anda,” *sotech.co.id*, 2020. <https://sotech.co.id/data-mining-tools/>.
- [4] P. C. H. Indonesia, “Apa Itu Visual Code Studio? Fitur dan Keunggulannya,” *idcloudhost.com*, 2022. <https://idcloudhost.com/panduan/visual-code-studio-adalah/>.
- [5] R. Setiawan, “Flowchart Adalah: Fungsi, Jenis, Simbol, dan Contohnya,” *repository.undar.ac.id*, 2021. <https://www.dicoding.com/blog/flowchart-adalah/>.
- [6] S. subekti Wasis, “Perbedaan PHP Native dan Framework,” *medium.com*, 2019. <https://medium.com/@sigitwasissubekti/perbedaan-php-native-dan-framework-18424a3b87ee> (accessed Dec. 25, 2021).
- [7] Waryanto, “Pengertian Website Lengkap dengan Jenis dan Manfaatnya,” *repository.itelkom-pwt.ac.id*, 2021. <https://www.niagahoster.co.id/blog/pengertian-website/> (accessed Dec. 25, 2021).
- [8] Salamadian, “BASIS DATA: Pengertian, Komponen dan Sistem Basis Data (Database),” *salamadian.com*, 2018. <https://salamadian.com/pengertian-basis-data-database/>.
- [9] L. Azizah, “Apa itu Database? Jenis, Fungsi dan Manfaatnya,” *gramedia.com*, 2021. <https://www.gramedia.com/best-seller/apa-itu-database/>.
- [10] Y. Mardi, “Data Mining: Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatika,” *jim.unindra.ac.id*, vol. 2, no. 2, pp. 213–219, 2019.