

**KLASIFIKASI KELAYAKAN PENERIMA BANTUAN BERAS MISKIN
MENGUNAKAN METODE KNN (K-NEAREST NEIGHBOR) DAN NAIVE BAYES
STUDI KASUS: DESA DRINGO, KECAMATAN TODANAN,
KABUPATEN BLORA**

Siti Nawala Jilan¹, Fandy Indra Pratama²

^{1,2} Program Studi Teknik Informatika, Fakultas Teknik, Universitas Wahid Hasyim
Jl. Raya Gunungpati No.KM.15, Nongkosawit, Gunungpati, Semarang 50224
Email: sitinawalajilan11@gmail.com

Abstrak

Beras miskin (raskin) adalah program bantuan yang dikeluarkan oleh pemerintah dan diperuntukkan untuk masyarakat yang mempunyai kebutuhan kurang dan penghasilan yang rendah. Aktivitas penyaluran beras miskin di Desa Dringo diperuntukkan bagi keluarga yang tidak mampu maupun belum berkecukupan, tetapi sangat jauh menurut apa yang sudah diharapkan, sekarang keluarga yang berkecukupan dan penghasilan tinggi juga menerimanya. Hal ini merupakan sesuatu yang membutuhkan sebuah metode algoritma untuk bisa mencegah kesalahan dalam memberikan kelayakan untuk menerimanya. Penelitian ini bertujuan untuk melakukan prediksi kelayakan penerima beras miskin dengan data yang diperoleh dari Desa Dringo. Penelitian-penelitian terdahulu sudah dilakukan namun belum diketahui metode mana yang terbaik pada klasifikasi kelayakan penerima beras miskin. Dalam penelitian ini dilakukan komparasi metode Naïve Bayes dan Metode KNN (K-Nearest Neighbor). Dengan menggunakan metode Naïve bayes didapatkan akurasi sebesar 85.53% dan dengan menggunakan metode KNN (K-Nearest Neighbor) didapatkan nilai akurasi sebesar 93.42 %. Sesudah dikomparasi hasil tadi memperlihatkan bahwa dengan memakai metode KNN (K-Nearest Neighbor) dihasilkan akurasi tertinggi. Hal tadi menyimpulkan bahwa metode K-nearest neighbor mempunyai kinerja lebih baik dibandingkan menggunakan metode naïve bayes.

Kata kunci: Klasifikasi, Beras Miskin, Metode KNN dan Naive Bayes

1. PENDAHULUAN

Beras miskin (raskin) adalah menjadi subsidi pangan utama pada bentuk beras yang dikhususkan untuk keluarga miskin menjadi upaya berdasarkan pemerintah buat mempertinggi ketahanan pangan dan menaruh proteksi dalam keluarga miskin (Tambunan et al., 2019). Permasalahan yang selalu terjadi dalam penyaluran donasi beras miskin umumnya masih memakai asumsi saja dan belum ada sebuah prediksi dalam waktu penyaluran pemisahan penerima beras miskin tersebut. Kemudian kecil atau besarnya masyarakat terkadang protes karena warga yang seharusnya mendapatkan bantuan tetapi mereka tidak mendapatkan bantuan (Suryeni et al., 2015).

Raskin adalah subsidi pangan pada bentuk beras yang diberikan untuk rumah tangga berpenghasilan rendah menjadi upaya menurut pemerintah buat menaikkan ketahanan pangan dan menaruh proteksi sosial dalam rumah tangga sasaran. Tujuan Program Beras buat Rakyat Miskin (Raskin) program Raskin adalah subsidi pangan menjadi upaya menurut Pemerintah dalam menaikkan ketahanan pangan dan menaruh proteksi dalam keluarga miskin melalui pendistribusian beras yang dibutuhkan bisa menjangkau keluarga miskin. Sasarannya merupakan terbantu dan terbukanya akses keluarga miskin yang sudah terdata menggunakan kuantum tertentu cocok dengan hasil musyawarah desa/kelurahan menggunakan harga bersubsidi pada tempat, sebagai akibatnya bisa membantu menaikkan ketahanan pangan keluarga miskin. Pokok pengelolaan Raskin merupakan suatu nilai-nilai dasar yang selalu sebagai landasan atau acuan pada setiap pengambilan keputusan juga tindakan yang akan diambil pada pelaksanaan rangkaian aktivitas

Raskin. Nilai-nilai dasar tadi diyakini sanggup mendorong terciptanya tujuan Raskin (Ridwan & Rahman, 2019).

Desa Dringo adalah satu diantara desa yang berada di Kecamatan Todanan, dan beralamat di Jalan Raya Todanan – Japah KM. 03 Ds. Dringo – Todanan – Blora. Desa Dringo memiliki luas daerah sebesar 365 Ha, secara umum topografi Desa Dringo juga mempunyai 2 Dusun. Jumlah penduduk Desa Dringo terdiri dari 761 kepala keluarga dan 2317 jumlah penduduknya (Data monografi desa).

Aktivitas penyaluran pembagian beras miskin pada Desa Dringo yang diperuntukkan kepada keluarga yang tidak layak maupun belum berkecukupan pastinya sangat jauh dari apa yang sudah diharapkan, sebab itu beras miskin yang semula diperuntukkan buat keluarga yang tidak layak, kini keluarga yang berkecukupan dan penghasilan besar juga menerimanya. Bahkan ada juga keluarga yang tidak layak tidak mendapatkan donasi sedangkan keluarga yang berkecukupan mendapatkan beras miskin tersebut. Hal ini terjadi disebabkan tidak adanya sebuah metode yang dipakai sebagai akibatnya pembagian bantuan tidak tertata dengan baik. Maka dengan itu sangat perlu adanya sebuah prediksi atau sebuah algoritma supaya bisa mencegah kesalahan dalam penyaluran pembagian beras miskin

Berdasarkan latar belakang tersebut, maka dapat dirumuskan suatu permasalahan yaitu bagaimana cara menghasilkan sistem prototype untuk klasifikasi kelayakan penerima bantuan beras miskin pada desa Dringo dengan hasil yang akurat? Batasan masalah yang akan dibahas dalam penelitian ini yaitu data yang digunakan untuk penelitian merupakan data yang berasal dari kelurahan Desa Dringo tahun 2020-2021. *Output* yang dapat dihasilkan terdiri dari layak dan tidak layak mendapatkan bantuan beras miskin. *Prototype* sistem menggunakan aplikasi Matlab. Tujuan penelitian ini, yaitu menghasilkan sistem *prototype* klasifikasi kelayakan penerima bantuan beras miskin pada desa Dringo berdasarkan tingginya tingkat akurasi antara metode KNN (*K-Nearest Neighbor*) dan *naïve bayes*. Selain itu untuk mengetahui output penerapan data *mining* pada klasifikasi penerima bantuan beras miskin dalam masyarakat desa, yang nantinya bisa bermanfaat bagi lembaga kemasyarakatan untuk menyalurkan bantuan supaya lebih tepat sasaran.

Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode KNN (*K-Nearest Neighbor*) (Hasanah et al., 2019). Dalam penelitiannya tersebut peneliti melakukan pengujian *K-Fold Cross Validation* dalam algoritma *K-Nearest Neighbor* untuk memprediksi penerimaan dana bantuan desa. Metode yang digunakan yaitu KNN (*K-Nearest Neighbor*) metode klasifikasi dengan mencari jeda terdekat antara data yang akan dinilai menggunakan K tetangga (*neighbor*) terdekatnya pada data pelatihan. Berdasarkan penelitian yang dilakukan dihasilkan kesimpulan bahwa, klasifikasi terhadap *dataset* penerimaan bantuan dana desa dilakukan melalui tahapan perhitungan manual buat mengetahui *output* kelayakan maupun ketidaklayakan menurut data baru, serta penggunaan *software Rapidminer* untuk menguji akurasi *dataset* pada bermacam nilai K. Hasilnya membuktikan maka dengan K=15 dan K=30 data baru (D160) memiliki kategori “Tidak Layak” dengan tingkat akurasi sebesar 100%. Kemudian dengan K=45, K=60 dan K=75 data baru (D160) memiliki kategori “Layak” dengan tingkat akurasi sebesar 81,25%.

Beras Miskin (raskin) adalah rencana bantuan yang diberikan dari pemerintah dan diperuntukkan untuk warga yang memiliki kebutuhan menurun dan memiliki penghasilan sedikit (Manik et al., 2015). Sasarannya utamanya diperuntukkan bagi warga miskin memiliki tujuan dalam membantu untuk penyaluran bantuan penerimaan beras miskin, pada penyalurannya bagi warga yang telah tercatat melalui metode kuantum variabel yang telah dipenuhi oleh desa bahwa akan menerima bantuan (Ridwan & Rahman, 2019).

Data *mining* sangat diperlukan bagi dunia kerja dan bisa membantu perusahaan untuk menerima pola fakta menurut data-data yang sudah tercatat sebelumnya. *Knowledge* tentang data *mining* yang didapatkan tadi akan sebagai panduan untuk menarik upaya pemeliharaan dan peningkatan pada taraf usaha sebuah perusahaan (Abdillah et al., 2016). Data *mining* merupakan serangkaian metode untuk menemukan pola-pola data. Metode ini otomatis maupun seringnya semi otomatis. Pola yang didapatkan wajib memiliki arti dan sanggup menaruh sebuah keuntungan, dan

keuntungannya sanggup secara keuangan dan data yang diperlukan berskala besar (Santoso et al., 2016).

Klasifikasi merupakan metode yang digunakan untuk mengumpulkan suatu himpunan model atau fungsi. Klasifikasi dapat digunakan untuk mendeskripsikan dan membedakan kelas-kelas data maupun konsep-konsep. Tujuan dari klasifikasi yaitu bisa memakai contoh tadi buat memprediksi kelas berdasarkan suatu objek yang mana kelasnya belum didapatkan dalam membedakan kelas-kelas data, sehingga model tersebut bisa dipakai dalam memprediksi nilai suatu kelas yang belum diketahui (Rani, 2016). Klasifikasi adalah membedakan konsep berdasarkan model atau fungsi yang ditemukan dengan tujuan agar objek yang tidak diketahui labelnya dapat diperkirakan kelasnya. Ketika melaksanakan tujuan tersebut, metode klasifikasi membangun suatu contoh menurut aturan atau fungsi yang sudah ditentukan, dimana model tersebut dapat membedakan data di dalam kelas-kelas yang berbeda. Model itu sendiri bisa berupa aturan “jika-maka”, berupa pohon keputusan, atau formula *matredies* (Bustami, 2010). Klasifikasi data merupakan proses mencari kesamaan properti yang terdapat pada himpunan obyek di dalam sebuah basis data kemudian mengklasifikasikannya di dalam kelas-kelas yang berbeda menurut contoh klasifikasi yang disetujui. Klasifikasi bertujuan untuk menemukan model pada *training set* yang melainkan atribut di dalam kelas yang sesuai, kemudian model tersebut dipakai dalam mengklasifikasikan atribut yang belum diketahui kelasnya. Cara klasifikasi dibagi menjadi beberapa teknik, salah satunya merupakan Pohon Keputusan (Ginting et al., 2014). Berdasarkan penjelasan tersebut, maka dapat disimpulkan bahwa klasifikasi adalah proses pengelompokan data berdasarkan kelas-kelas tertentu dari suatu objek.

Naïve Bayes Classifier merupakan metode klasifikasi yang berdasarkan *teorema Bayes*. Metode klasifikasi ini memakai probabilitas & statistik yang pertama kali diketemukan sang ilmuwan bangsawan Inggris bernama Thomas Bayes, yakni suatu metode buat memprediksi peluang pada masa yang akan tiba dari peristiwa pada masa sebelumnya, sebagai akibatnya metode ini dikenal menjadi Teorema Bayes. Atribut utama dari *Naive Bayes Classifier* ini merupakan asumsi yang paling meyakinkan (Syukri Mustafa et al., 2017).

Rumus dari perhitungan *Naive Bayes* dalam buku berjudul *Data Mining Concept and Techniques*, Nirmala dalam (Han et al., 2011) yaitu :

$$P = (C_i | X = p \quad (1)$$

Atau ditulis dengan:

$$Posterior = \frac{\text{likelihood} \times \text{priorprobability}}{\text{evidence}} \quad (2)$$

Keterangan:

$P(C_i|X)$ = Peluang kategori C_i jika diberikan fakta atau bukti X (*posterior*)

$P(X|C_i)$ = Peluang pada kategori C_i , dimana fakta atau bukti X muncul menurut tingkatan tersebut (*likelihood*)

$P(C_i)$ = Peluang dari kategori yang diberikan, disamakan dengan tingkatan lainnya yang dianalisa (*prior probability*)

$P(X)$ = Total peluang dari fakta maupun bukti X (*evidence*)

Split Validation merupakan teknik validasi dengan cara membuat data terbagi menjadi dua bagian secara acak. Dimana data tersebut dibagi menjadi data *training* serta data *testing*. Data training ialah data yang digunakan sebagai data awal percobaan untuk model data mining yang baik. Data testing ialah data yang akan digunakan sebagai pengujian dari model yang telah dibuat. Dengan menggunakan *split validation* maka akan dilakukan percobaan pada training berdasarkan pada *split ratio* yang telah ditentukan. Kemudian sisa dari *split ratio* pada data training digunakan

untuk data testing. Misalkan jika *split ratio* pada data training 90 % maka data testing sebesar 10 % (Fajri, 2018).

RapidMiner adalah aplikasi data *mining* berbasis *open source* yang terkemuka dan ternama. Didalamnya terdapat aplikasi yang berdiri sendiri untuk analisis data dan sebagai mesin data *mining* seperti untuk *loading* data, transformasi data, pemodelan data, dan metode visualisasi data. *RapidMiner* pertama kali dinamai *Yet Another Learning Environment* atau disingkat YALE. Pada tahun 2007 akhirnya diganti namanya menjadi *RapidMiner* (Nofitri & Irawati, 2019)

Algoritma *K-Nearest Neighbor* merupakan salah satu metode yang dipakai untuk pengklasifikasian. Prinsip kerja *K-Nearest Neighbor* (KNN) merupakan pencarian jeda terdekat dari data yang untuk dievaluasi dengan *K* tetangga (*neighbor*) terdekatnya dalam data pelatihan. Teknik ini termasuk dalam kelompok klasifikasi *non parametric*. Algoritma *KNN* menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari data uji yang baru. Jarak yang digunakan adalah jarak *Euclidean Distance*. Jarak *Euclidean* adalah jarak yang paling umum digunakan pada data numerik. Nilai *k* yang terbaik untuk algoritma ini tergantung pada data. Secara umum, nilai *k* yang tinggi akan mengurangi efek *noise* pada klasifikasi. Kasus spesifik dimana klasifikasi terstruktur mengenai *training* data yang paling dekat disebut algoritma *K-Nearest Neighbor* (Ula & Fuadi, 2016).

Algoritma *K-Nearest Neighbor* sebagai berikut:

1. Tentukan parameter *K*
2. Hitung jarak antara data yang untuk dievaluasi dengan semua pelatihan
3. Urutkan jarak yang terbentuk (dari terkecil ke terbesar)
4. Tentukan jarak terdekat sejumlah *K*
5. Pasangkan kelas yang bersesuaian
6. Cari total kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

Rumus KNN:

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2} \quad (3)$$

Keterangan:

x_1 = Data Uji
 x_2 = Data Sampel
 i = Variabel Data
 d = Jarak
 p = Dimensi Data

Confusion Matrix merupakan alat yang digunakan untuk menganalisis bagaimana pengklasifikasi (*Classifier*) dapat mengenali *tuple-tuple* pada kelas-kelas yang berbeda. (Hermawanti, 2012).

Dalam kasus dengan dua klasifikasi data keluaran, seperti contoh “C1” dan “C2”, atau contoh lainnya tiap kelas yang diprediksi memiliki empat kemungkinan keluaran yang berbeda, yaitu *true positif* (TP), *true negative* (TN), *false positif* (FP), *false negative* (FN) menunjukkan ketepatan klasifikasi dapat dilihat di Tabel 1.

Tabel 1. Ilustrasi Akurasi Matrix

Kelas	Terklasifikasi Positive	Terklasifikasi Negative
Positif	True Positive (TP)	False Negative (FN)
Negatif	False Positive (FP)	True Negative (TN)

Nilai akurasi ini menggambarkan seberapa akurat *system* dapat mengklasifikasi data secara benar, nilai akurasi adalah persamaan antara data yang terklasifikasi benar beserta keseluruhan data.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (4)$$

$$Presisi = \frac{TP}{FP+TP} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{FN+TP} \times 100\% \quad (6)$$

Keterangan:

- TP merupakan *True Positive*, yakni jumlah data positif yang terklasifikasi dengan benar oleh sistem.
- TN ialah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FN adalah *False Negative*, yakni jumlah data negatif tetapi terklasifikasi salah oleh sistem.
- FP ialah *False Positive*, yakni jumlah data positif tetapi terklasifikasi salah oleh sistem.
- I* merupakan jumlah kelas,

Matlab merupakan perangkat lunak yang digunakan untuk pemrograman, analisis, serta komputasi teknik dan matematis berbasis *matriks*. Matlab merupakan singkatan dari Matrix Laboratory lantaran bisa menyelesaikan kasus perhitungan pada bentuk *matriks*. Matlab versi pertama dirilis dalam tahun 1970 sang Cleve Moler. Pada awalnya, Matlab di desain buat menuntaskan perkara-perkara persamaan aljabar linear. Seiring berjalannya waktu, acara ini terus mengalami perkembangan menurut segi fungsi & performa komputasi. Matlab bisa dioperasikan dalam sistem operasi *windows*, *Linux*, juga *MacOS*. Selain itu, Matlab pula mampu dihubungkan menggunakan *software* maupun bahasa pemrograman eksternal lainnya, seperti *C*, *Java*, *NET*, dan *Microsoft Excel*. Dalam matlab tersedia juga kotak kakas (*toolbox*) yang bisa dipakai buat aplikasi-aplikasi khusus, contohnya pengolahan sinyal, sistem kontrol, logika *fuzzy*, jaringan saraf tiruan, optimasi, pengolahan citra digital, bioinformatika, simulasi, dan berbagai teknologi lainnya (Tjolleng, 2017).

2. METODE PENELITIAN

Penelitian ini, peneliti melakukan observasi ke salah satu Desa Dringo di Kabupaten Blora, dengan tujuan untuk mengumpulkan data-data yang dibutuhkan untuk penelitian ini. Data-data yang dibutuhkan yakni data berkaitan dalam penerimaan bantuan beras miskin (raskin). Masalah-masalah yang telah dihimpun dari hasil observasi, peneliti melakukan pencarian dari berbagai macam sumber dokumen yang mendukung dalam penelitian ini. Diantaranya buku, artikel, jurnal dan *website* internet yang menyimpan berita dan terbukti kebenarannya dengan permasalahan yang sedang peneliti teliti.

Pengumpulan data diawali dengan datang ke kantor kelurahan Desa Dringo dan melakukan wawancara untuk mendapatkan data daftar penerima beras miskin (Raskin). Data yang diperoleh berbentuk data mentah data penerima beras miskin yang berjumlah 761 record. Data yang didapatkan wajib dengan tahap pemilihan dan *preproccesing* sebelum dipakai untuk penelitian antara lain. Berikut adalah atribut dari data penerima raskin dapat dilihat di Tabel 2.

Tabel 1. Atribut Data Raskin

No	Nama Atribut	Keterangan
1	Nama	Nama yang layak menerima bantuan beras miskin
2	Tempat Lahir	Tempat lahir data penduduk desa dringo
3	Alamat	Alamat RT dan RW
4	Kelurahan	Kelurahan yang di tempati
5	Kondisi Rumah	Kondisi rumah (Permanen /semi permanen)
6	Jumlah Penghasilan	Penghasilan setiap bulannya
7	Penerangan Yang Digunakan	Penerangan yang digunakan di rumah
8	Pekerjaan	Pekerjaan yang dijalankan
9	Status	Layak atau Tidak layaknya mendapatkan bantuan

Berdasarkan penelitian terkait yang menggunakan atribut-atribut yang berkaitan dengan keterangan penerima beras miskin yang berasal dari kelurahan Desa Dringo Kecamatan Todanan. Seluruh data pada atribut yang terpilih tersebut akan digabungkan untuk memperoleh hasil kelayakan penerima beras miskin dalam bentuk klasifikasi kategori. Pada *preprocessing* terdiri dari dua tahap, yaitu

a. Pengolahan Data

Dari data yang dihasilkan kemudian dilakukan *preprocessing* dengan merapikan tulisan yang tidak rapi atau salah tulis.

b. *Sortir*

Dari hasil pengolahan data kemudian akan dibagi menjadi 2 merupakan data *Training* dan data *Testing* yang mana data *training* berjumlah 685 (90%) data dan data *Testing* berjumlah 76 (10%) data.

Proses perhitungan dalam sistem dengan menggunakan metode KNN adalah algoritma klasifikasi yang mencari jarak terdekat dari data yang buat dievaluasi pada tetangga data terdekatnya pada data pelatihan.

Setelah dataset tertera selanjutnya mengimplementasikan metode. Disini untuk mengimplementasikannya ada 2 yaitu:

1. Menghitung nilai kondisi probabilitas, Dimana tahap ini menghitung nilai C1 dan C2 total yang nantinya akan mendapatkan model dan model tersebut nanti akan dipakai untuk data testing, juga menghitung C1 dan C2 dari parameter masing-masing atribut.
2. Menghitung nilai prediksi untuk *output class*, dimana tahap ini untuk menghitung status dari tiap-tiap data perorang yang telah terdata yang dilakukan pada data *testing*.

Setelah melakukan perhitungan hasil dari klasifikasi data yang sudah dihitung dengan metode KNN (*K-Nearest Neighbor*) dan Metode *Naive Bayes*, selanjutnya menghitung tingkat akurasi dan presisi pada setiap data warga yang sesuai dengan atribut dan parameter sehingga nanti dapat diketahui nilai akurasi dan presisi untuk membandingkan nilai asli dengan dataset.

Setelah melakukan perhitungan akurasi dengan memakai metode KNN (*K-Nearest Neighbor*) dan *Naive Bayes* maka akurasi yang dihasilkan dari kedua metode tersebut kemudian dikomparasi untuk didapatkannya nilai akurasi yang tertinggi.

3. HASIL DAN PEMBAHASAN

Pada tahap awal dari penelitian ini adalah pengumpulan data yang nantinya akan dilakukan *preprocessing* untuk mendapatkan data yang sesuai. Data yang digunakan merupakan data dari kelurahan Desa Dringo Kecamatan Todanan dari tahun 2020 sampai 2021 yang didapatkan dari kelurahan yang berjumlah 761 *record*. Data tersebut didapatkan berdasarkan permintaan dari pemohon yang menyertakan atribut apa saja yang hendak diminta dan pihak kelurahan data akan memberikan data berdasarkan permintaan pemohon. Atribut pada data tersebut adalah sesuai dengan Tabel 3.

Tabel 3. Daftar Atribut

No	Atribut	Keterangan
1	Kondisi Rumah	Kondisi rumah (Permanen/semi permanen)
2	Jumlah Penghasilan	Penghasilan setiap bulannya
3	Penerangan Yang Digunakan	Penerangan yang digunakan di rumah
4	Pekerjaan	Pekerjaan yang dijalankan

Atribut yang digunakan berdasarkan penelitian ini memiliki 4 atribut, untuk pembagian datanya jumlah data training sebanyak 685 *record*, dan data testing sebanyak 76 *record*. Sehingga total dataset yang digunakan berjumlah 761 *record*.

Berikut adalah 4 sampel dari data *training* yang terkumpul sesuai dengan Tabel 4.

Tabel 4. Data Training

Kondisi Rumah	Jumlah Penghasilan	Penerangan	Pekerjaan	Kategori
Papan	Sedang	Listrik	Buruh	Tdk Layak
Papan	Sedang	Listrik	Petani	Layak
Tembok	Sedang	Listrik	Petani	Tidak Layak
Papan	Sedang	Menyambung	Buruh	Layak

Dari Tabel 4 dapat dijelaskan nilai perumpamaannya sebagai berikut:

1. Kondisi Rumah dengan kriteria Papan memiliki nilai “1” dan Tembok memiliki nilai “2”
2. Jumlah Penghasilan dengan kriteria Rendah memiliki nilai “1”, Sedang memiliki nilai “2” dan Tinggi memiliki nilai “3”
3. Penerangan yang digunakan dengan kriteria Menyambung memiliki nilai “1” dan Listrik memiliki nilai “2”
4. Pekerjaan dengan kriteria Buruh memiliki nilai “1”, Pedagang memiliki nilai “2”, Petani memiliki nilai “3” dan Wirausaha memiliki nilai “4”
5. Kategori dengan kriteria Layak memiliki nilai “1” dan Tidak Layak memiliki nilai “2”

Berdasarkan Tabel 4 maka dapat diberikan keterangan sebagai berikut:

a. Kondisi Rumah

Merupakan variabel yang menjelaskan tentang suatu kondisi rumah, kemudian dikelompokkan membentuk 2 bagian merupakan: Papan dan Tembok.

Tabel 5. Kondisi Rumah

No	Kondisi Rumah
1	Papan
2	Tembok

Sumber: Data Desa Dringo 2020

b. Jumlah Penghasilan

Merupakan variabel yang dikelompokkan membentuk 3 bagian ialah: Rendah, Sedang dan Tinggi. Kriteria tersebut bisa dilihat pada Tabel 6.

Tabel 6. Jumlah Penghasilan

No	Jumlah Penghasilan	Keterangan
1	Rendah	<Rp.1.000.000
2	Sedang	Rp.1.000.000 – 2.500.000
3	Tinggi	>Rp.2.500.000

Sumber: Data Desa Dringo 2020

c. Penerangan yang digunakan

Merupakan variabel yang dikelompokkan membentuk 2 bagian merupakan: Menyambung dan Listrik bisa dilihat pada Tabel 7.

Tabel 7. Penerangan yang digunakan

No	Penerangan
1	Menyambung
2	Listrik

Sumber: Data Desa Dringo 2020

d. Pekerjaan

Merupakan variabel yang dikelompokkan membentuk 4 bagian yaitu Pedagang, Wirausaha, Buruh dan Petani dapat dilihat pada Tabel 8.

Tabel 8. Kriteria Pekerjaan

No	Pekerjaan
1	Buruh
2	Pedagang
3	Petani
4	Wirausaha

Sumber: Data Desa Dringo 2020

Berikut adalah 3 sampel dari data *testing* yang terkumpul bisa dilihat pada Tabel 9.

Tabel 9. Data Testing

Kondisi Rumah	Jumlah Penghasilan	Penerangan	Pekerjaan	Kategori
Papan	Rendah	Menyambung	Buruh	Layak
Tembok	Rendah	Menyambung	Petani	Tidak Layak
Papan	Rendah	Menyambung	Pedagang	Tidak Layak

Sumber: Data Desa Dringo 2020

Dari Tabel 9 dapat dijelaskan nilai perumpamaannya sebagai berikut :

1. Kondisi Rumah dengan kriteria Papan memiliki nilai “1” dan Tembok memiliki nilai “2”
2. Jumlah Penghasilan dengan kriteria Rendah memiliki nilai “1”, Sedang memiliki nilai “2” dan Tinggi memiliki nilai “3”
3. Penerangan yang digunakan dengan kriteria Menyambung memiliki nilai “1” dan Listrik memiliki nilai “2”
4. Pekerjaan dengan kriteria Buruh memiliki nilai “1”, Pedagang memiliki nilai “2”, Petani memiliki nilai “3” dan Wirausaha memiliki nilai “4”
5. Kategori dengan kriteria Layak memiliki nilai “1” dan Tidak Layak memiliki nilai “2”

Tabel 9 merupakan tabel data *testing* yang akan dilakukan sebuah perhitungan manual dengan menggunakan rumus algoritma KNN (*K-Nearest Neighbor*) sebagai berikut.

Berikut ini merupakan contoh perhitungan manual untuk mencari nilai *euclidean distance* antara data training) ke data testing.

$$d_i = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Data Testing 1:

Kondisi Rumah Papan nilainya =1

Jumlah Penghasilan Rendah nilainya = 1

Penerangan Yang Digunakan Menyambung nilainya = 1

Pekerjaan Buruh nilainya = 1

Data Training 1:

Kondisi Rumah Papan nilainya =1

Jumlah Penghasilan Sedang nilainya = 2

Penerangan Yang Digunakan Listrik nilainya = 2
 Pekerjaan Pedagang nilainya = 2

D (Data Testing (1), Data Training:

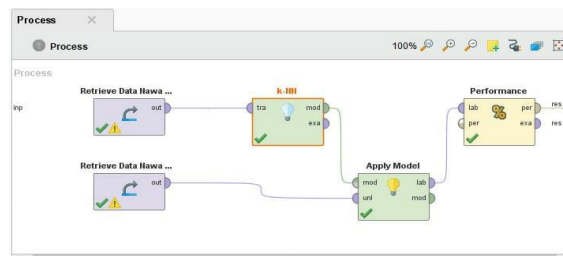
$$d(1,685) = \sqrt{(1-1)^2 + (1-2)^2 + (1-2)^2 + (1-1)^2}$$

$$d(1,685) = \sqrt{0 + 1 + 1 + 0}$$

$$d(1,685) = \sqrt{2}$$

$$d(1,685) = 1.414$$

Model pengujian ini dimana pada dataset source nantinya akan disesuaikan dengan dataset penerima beras miskin. Kemudian untuk memvalidasi model dari algoritma KNN (K-Nearest Neighbor) digunakan metode *split validation*. Dimana didalamnya terdapat *performance* dengan menggunakan *confussion matrix* sebagai model evaluasi dari kinerja algoritma KNN. Berikut adalah penerapan algoritma dengan menggunakan *software rapidminer*. Penerapan tersebut bisa dilihat pada Gambar 1.



Gambar 1. KNN Sub Proses Split Validation

Dalam sub proses *split validation* terdapat dua bagian dimana ada *training* dan ada *testing* bisa dilihat pada Gambar 1 pada bagian *training* terdapat algoritma KNN hal tersebut dimaksudkan agar dataset dibuat modelnya menggunakan algoritma KNN dimana pada *split validation* dengan nilai $k=10$ *folds* sehingga terbagi data menjadi 10: 90 yakni 10 % dijadikan sebagai data *testing* dan 90% dijadikan sebagai data *training*. Kemudian pada data *testing* terdapat dua fitur yaitu *apply model* yang digunakan untuk menerapkan model data yang pernah dibentuk sebelumnya dengan data *testing*.

Terakhir dibagian testing terdapat fitur *performance* dimana fitur tersebut digunakan untuk mengevaluasi hasil kinerja algoritma KNN dengan parameter pengukuran *confussion matrix* untuk mengetahui nilai akurasi. Hasilnya bisa dilihat pada Gambar 2. Data *training* yang dipakai dalam penelitian ini berjumlah 685 *record* dan data *testing* berjumlah 76 *record*.

accuracy: 93.42%			
	true 1	true 2	class precision
pred 1	53	4	92.98%
pred 2	1	18	94.74%
class recall	98.15%	81.82%	

Gambar 2. Confussion Matrix Model pada Aplikasi Rapidminer

Dari Gambar 2. bisa dilihat hasil *performance* yang dihasilkan menggunakan software *RapidMiner Performance* vektor Accuracy: 93.42 %.

Setelah diketahui hasil analisa antara data *training* dan data *testing* pada *Rapidminer* dengan menggunakan algoritma KNN dengan total data *testing* berjumlah 76 *record*. Maka dapat dihitung untuk mencari nilai akurasi dengan menggunakan *confussion matriks* dalam perhitungan manual seperti yang ditunjukkan pada Tabel 10.

Tabel 10. Confusion Matriks

	True Layak	True Tdk Layak
Pred. Layak	53	4
Pred.Tdk Layak	1	18

Berdasarkan Tabel 10. yang menunjukkan jumlah sampel yang bernilai benar, sehingga dapat dihitung nilai akurasi sebagai berikut:

$$\text{Akurasi} = \frac{\text{Jumlah Data Yang Benar}}{\text{Total Data}} \times 100\%$$

$$\text{Akurasi} = \frac{53 + 1}{53 + 4 + 1 + 18} \times 100\%$$

$$\text{Akurasi} = \frac{53 + 18}{76} \times 100\%$$

$$\text{Akurasi} = \frac{71}{76} \times 100\%$$

$$\text{Akurasi} = 0,9342 \times 100\%$$

$$\text{Akurasi} = 93.42 \%$$

Berdasarkan hasil perhitungan tersebut maka diketahui nilai akurasi yang dihasilkan dari 76 data testing penerima bantuan beras miskin sebesar 93.42 %.

Dilakukannya sebuah perhitungan dari data testing dengan perhitungan manual menggunakan rumus algoritma *naive bayes* sebagai berikut.

$P(C_i)$

$P(\text{STATUS KELAYAKAN} = \text{"LAYAK"}) = 507/685$

$P(\text{STATUS KELAYAKAN} = \text{"TIDAK LAYAK"}) = 178/685$

Perhitungan data testing (1,1):

$P(X|C_i)$

$P(\text{Kondisi Rumah} = \text{"Papan"} | \text{Status Kelayakan} = \text{"Layak"}) = 436/507$

$P(\text{Kondisi Rumah} = \text{"Papan"} | \text{Status Kelayakan} = \text{"Tidak Layak"}) = 110/507$

$P(\text{Jumlah Penghasilan} = \text{"Sedang"} | \text{Status Kelayakan} = \text{"Layak"}) = 190/507$

$P(\text{Jumlah Penghasilan} = \text{"Sedang"} | \text{Status Kelayakan} = \text{"Tidak Layak"}) = 94/178$

$P(\text{Penerangan} = \text{"Listrik"} | \text{Status Kelayakan} = \text{"Layak"}) = 183/507$

$P(\text{Penerangan} = \text{"Listrik"} | \text{Status Kelayakan} = \text{"Tidak Layak"}) = 95/178$

$P(\text{Pekerjaan} = \text{"Buruh"} | \text{Status Kelayakan} = \text{"Layak"}) = 237/507$

$P(\text{Pekerjaan} = \text{"Buruh"} | \text{Status Kelayakan} = \text{"Tidak Layak"}) = 19/178$

$P(X|\text{Status Kelayakan} = \text{"Layak"}) = P(\text{Kondisi Rumah} = \text{"Papan"}, \text{Jumlah Penghasilan} = \text{"Sedang"}, \text{Penerangan} = \text{"Listrik"}, \text{Pekerjaan} = \text{"Buruh"} | \text{Status Kelayakan} = \text{"Layak"})$

$= 436/507 * 190/507 * 183/507 * 237/507$

$= 0.859 * 0.374 * 0.360 * 0.647$

$= 0.0540$

$P(X|\text{Status Kelayakan} = \text{"Tidak Layak"}) = P(\text{Kondisi Rumah}$

$= \text{"Papan"}, \text{Jumlah Penghasilan} = \text{"Sedang"}, \text{Penerangan} = \text{"Listrik"}, \text{Pekerjaan} = \text{"Buruh"} | \text{Status Kelayakan} = \text{"Tidak Layak"})$

$= 110/178 * 94/178 * 95/178 * 19/178$

$= 0.617 * 0.528 * 0.533 * 0.106$

$= 0.0184$

$P(X|C_i) * P = (C_i)$

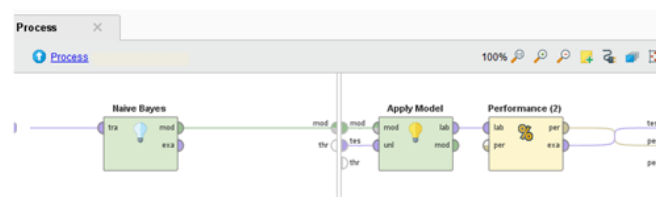
$P(X|\text{Status Kelayakan} = \text{"Layak"}) * P(\text{Status Kelayakan} = \text{"Layak"})$

$= 0.0540 * 507/685 = 0.039967$

$$P(X|\text{Status Kelayakan} = \text{"Tidak Layak"}) * P(\text{Status Kelayakan} = \text{"Tidak Layak"}) \\ = 0.0184 * 178/685 = 0.004781$$

Untuk kondisi rumah = "Papan", Jumlah penghasilan = "Sedang", Penerangan = "Listrik", Pekerjaan = "Buruh" Masuk ke kelas status kelayakan = "Layak" karena status kelayakan yang layak lebih besar dari status kelayakan yang tidak layak.

Model pengujian ini dimana pada dataset *source* nantinya akan disesuaikan dengan dataset penerima beras miskin. Kemudian untuk memvalidasi model dari algoritma *Naive Bayes* digunakan metode *split validation*. Dimana didalamnya terdapat *performance* dengan menggunakan *confussion matrix* sebagai model evaluasi dari kinerja algoritma *Naive Bayes*. Berikut adalah penerapan algoritma dengan menggunakan *software rapidminer*. Penerapan tersebut bisa dilihat pada Gambar 3.



Gambar 3. Naive Bayes Sub Proses Split Validation

Dalam sub proses *split validation* terdapat dua bagian dimana ada *training* dan ada *testing* bisa dilihat pada Gambar 3., pada bagian *training* terdapat algoritma *Naive Bayes* hal tersebut dimaksudkan agar dataset dibuat modelnya menentukan algoritma *Naive Bayes* dimana pada *split validation* dengan nilai $k=10$ folds sehingga memisah data menjadi 10:90 yakni 10 % dijadikan sebagai data *testing* dan 90% dijadikan sebagai data *training*. Kemudian pada data *testing* terdapat dua fitur yaitu *apply* model yang digunakan untuk menerapkan model data yang telah dibentuk sebelumnya dengan data *testing*.

Terakhir dibagian testing terdapat fitur *performance* dimana fitur tersebut digunakan untuk mengevaluasi hasil kinerja algoritma *Naive Bayes* dengan parameter pengukuran *confussion matrix* untuk mengetahui nilai akurasi. Hasilnya dapat dilihat pada Gambar 4. Data *training* yang dilakukan pada penelitian ini berjumlah 685 record dan data *testing* berjumlah 76 record.

accuracy: 85.53%

	true 1	true 2	class precision
pred. 1	54	11	83.08%
pred. 2	0	11	100.00%
class recall	100.00%	50.00%	

Gambar 4. Hasil Matrix yang di Rapidminer

Dari Gambar 4. bisa dilihat hasil performance yang dihasilkan menggunakan software RapidMiner Performance vektor Accuracy: 85.53%.

Setelah diketahui hasil analisa antara data *training* dan data *testing* pada *Rapidminer* dengan menggunakan algoritma *Naive Bayes* dengan total data *testing* sebanyak 76 record. Maka dapat dihitung akan mencari nilai akurasi dengan menggunakan *confussion matriks* dalam perhitungan manual.

Tabel 11. Confusion Matriks Naive Bayes

	True Layak	True Tdk Layak
Pred. Layak	54	11
Pred.Tdk Layak	0	11

Berdasarkan Tabel 11. yang menunjukkan jumlah sampel yang bernilai benar, sehingga dapat dihitung nilai akurasi sebagai berikut:

$$Akurasi = \frac{\text{Jumlah data yang benar}}{\text{Total data}} \times 100\%$$

$$Akurasi = \frac{54 + 11}{54 + 11 + 0 + 11} \times 100\%$$

$$Akurasi = \frac{65}{76} \times 100\%$$

$$Akurasi = 85.53\%$$

Berdasarkan hasil perhitungan tersebut maka diketahui nilai akurasi yang dihasilkan dari 76 data *testing* kelayakan beras miskin sebesar 85.53%.


Tabel 12. Komparasi Akurasi Metode KNN & Naive Bayes

Metode	Akurasi
KNN (K-Nearest Neighbor)	93.42%
Naive Bayes	85.53 %

Tabel 12 merupakan Komparasi metode KNN dan *Naive Bayes* dalam pengklasifikasian kelayakan penerima beras miskin. Pada Tabel 12 memperlihatkan dengan menggunakan metode KNN didapatkan nilai akurasi sebesar 93.42 % dan menggunakan metode *Naive Bayes* didapatkan nilai akurasi sebesar 85.53% .

Dari hasil komparasi tersebut menunjukkan metode KNN memiliki tingkat akurasi yang lebih tinggi. Hal tersebut menunjukkan bahwa kinerja metode KNN lebih bagus dibandingkan oleh metode *Naive Bayes*.

Berikut adalah *interface* rancangan *prototype* yang bisa dilihat pada Gambar 5.



Gambar 5. Perancangan Prototype

Setelah *prototype* aplikasi dapat dijalankan pada komputer, kemudian *user* melakukan simulasi untuk pengecekan *prototype* aplikasi yang telah terpasang tersebut dengan menginputkan data sebagai berikut:

1. Percobaan Prediksi Layak

Nama = Saidi
Kondisi Rumah = Papan
Penghasilan = Rendah
Penerangan = Menyambung
Pekerjaan = Buruh

Dari inputan tersebut dan klik Prediksi didapati hasil prediksi yaitu “Layak” seperti pada Gambar 6.

Gambar 6. Prediksi Layak

2. Percobaan Prediksi Tidak Layak

Nama = Minah
 Kondisi Rumah = Tembok
 Penghasilan = Rendah
 Penerangan = Menyambung
 Pekerjaan = Petani

Dari inputan tersebut dan klik Prediksi didapati hasil prediksi yaitu “Tidak Layak” seperti pada Gambar 7.

Gambar 7. Prediksi Tidak Layak

4. KESIMPULAN

Proses komparasi atau membandingkan antara metode *Naïve Bayes* dan metode KNN (*K-Nearest Neighbor*) dengan hasil pengujian yang dikerjakan model klasifikasi dengan menggunakan *Naïve Bayes* memiliki nilai optimal akurasi sebesar 85.53% dan menggunakan *split* rasio 10% sedangkan dengan menggunakan metode KNN memiliki nilai optimal akurasi sebesar 93.42% dengan *split* rasio 10%. Hal ini menunjukkan bahwa metode yang mempunyai nilai akurasi yang paling optimal adalah metode KNN (*K-Nearest Neighbor*) untuk klasifikasi kelayakan penerima bantuan beras miskin. Dengan adanya *prototype* system ini dengan algoritma ini dapat mengoptimalkan pembagian bantuan beras miskin agar tepat sasaran dan berjalan secara optimal. Untuk meningkatkan hasil kualitas penelitian ini agar dapat lebih baik lagi, diharapkan untuk penelitian selanjutnya dapat mengembangkan dan menambah atau merubah atribut sehingga mendapatkan hasil klasifikasi yang lebih baik lagi.

DAFTAR PUSTAKA

Abdillah, G., Putra, F. A., Renaldi, F., Informatika, P. S., Jenderal, U., Yani, A., Barat, J., &

- Cimahi, K. (2016). Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means. *Sentika 2016*, 2016(Sentika), 18–19.
- Bustami. (2010). Penerapan Algoritma Naive Bayes untuk Mengklasifikasi Data Nasabah. *TECHSI: Jurnal Penelitian Teknik Informatika*, 4, 127–146.
- Data Kelurahan Desa Dringo (2020). Profil Desa Dringo, Kecamatan Todanan, Kabupaten Blora.
- Fajri, F. N. (2018). *Perbandingan Sistem Klasifikasi Naïve Bayes dan Decision Tree Untuk Diagnosa Penyakit Diabetes*.
https://www.academia.edu/download/39122082/Perbandingan_Sistem_Klasifikasi_Naive_Bayes_dan_Decision_Tree_Untuk_Diagnosa_.pdf
- Ginting, S. L. B., Zarman, W., & Hamidah, I. (2014). *Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik*. November.
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. (3rd ed.).
- Hasanah, R. L., Hasan, M., Pangesti, W. E., Wati, F. F., & Gata, W. (2019). Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode KNN(K-Nearest Neighbor). *Jurnal Techno Nusa Mandiri*, 16(1), 1–6. <https://doi.org/10.33480/techno.v16i1.25>
- Hermawanti, L. (2012). Penerapan Algoritma Klasifikasi C4.5 untuk Diagnosis Penyakit Kanker Payudara. *Jurnal Teknik Unisfat*, 7(1), 57–64.
- Manik, A. R. S., Nurhadiyono, B., & Rahayu, Y. (2015). Implementasi Metode Weighted Product (WP) dalam Sistem Pendukung Keputusan untuk Menyeleksi Penerima Beras Masyarakat Miskin (Raskin). *Techno.COM*, 14(2), 109–114.
- Nofitri, R., & Irawati, N. (2019). Integrasi Metode Neive Bayes Dan Software Rapidminer Dalam Analisis Hasil Usaha Perusahaan Dagang. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, 6(1), 35–42. <https://doi.org/10.33330/jurteksi.v6i1.393>
- Rani, L. N. (2016). Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *INOVTEK Polbeng - Seri Informatika*, 1(2), 126. <https://doi.org/10.35314/isi.v1i2.131>
- Ridwan, D. A., & Rahman, B. (2019). Sistem Pendukung Keputusan Penentuan Penerimaan Raskin (Beras Miskin) Pada Kecamatan Kendari Barat Menggunakan Metode Analythical Hierarchi Process (Ahp). *Simtek : Jurnal Sistem Informasi Dan Teknik Komputer*, 4(1), 25–31. <https://doi.org/10.51876/simtek.v4i1.45>
- Santoso, H., Hariyadi, I. P., & Prayitno. (2016). Data Mining Analisa Pola Pembelian Produk Dengan Menggunakan Metode Algoritma Apriori. *Teknik Informatika*, 1, 19–24. <http://ojs.amikom.ac.id/index.php/semnasteknomedia/article/download/1267/1200>
- Suryeni, E., Agustin, Y. H., & Nurfitria, Y. (2015). Sistem Pendukung Keputusan Kelayakan Penerimaan Bantuan Beras Miskin Dengan Metode Weighted Product Di Kelurahan Karikil Kecamatan Mangkubumi Kota Tasikmalaya. *Konferensi Nasional Sistem & Informatika 2015*, 345–350.
- Syukri Mustafa, M., Rizky Ramadhan, M., & Thenata, A. P. (2017). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Citec Journal*, 4(2), 151–162.
- Tambunan, H. S., Sumarno, S., & Gunawan, I. (2019). Sistem Informasi Penyaluran Beras Miskin (Raskin) Kelurahan Desa Silau Malela Kabupaten Simalungun. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 1(1), 40. <https://doi.org/10.30865/json.v1i1.1378>
- Tjolleng, A. (2017). *Pengantar Pemrograman Matlab*. PT. Elex Media Komputindo, Jakarta
- Ula, M., & Fuadi. (2016). *Penentuan Klasifikasi Kelayakan Pembiayaan Murabahah Dalam Pemberian Modal Usaha Masyarakat Menggunakan K Nearest Neighbors [Determination of the Feasibility Classification of Murabahah Financing in Providing Community Business Capital Using K Nearest Nei. vol 4, 77–85.*