Jurnal Kajian dan Terapan Matematika
Volume 8, Edisi 2, Bulan Juli 2022, 75 - 84
http://journal.student.uny.ac.id/ojs/index.php/jktm

Klasifikasi penduduk miskin penerima PKH menggunakan metode naïve bayes dan KNN

Classification of poor residents recipients of PKH using naïve bayes and knnmethods

Essy Rahma Meilaniwati, Prodi Matematika FMIPA UNY
Drs. Muhammad Fauzan, M.Sc.St *, Prodi Matematika FMIPA UNY
*e-mail: mfauzan@uny.ac.id

Abstrak
Program Keluarga Harapan (PKH) merupakan salah satu program pemerintah sebagai upaya pengentasan kemiskinan penduduk yang telah terbukti dapat menanggulangi kemiskinan kronis di berbagai negara. Namun, banyak penduduk yang mengeluhkan kurang optimalnya penentuan penerima PKH, terutama di Kabupaten Bantul yang merupakan salah satu kabupaten di Yogyakarta dengan penduduk miskin terbanyak. Penelitian ini bertujuan untuk mengklasifikasi penduduk yang berhak menjadi penerima PKH menggunakan metode Naïve Bayes dan K-Nearest Neighbors (KNN). Metode Naïve Bayes memiliki konsep pengklasifikasian berdasarkan probabilitas tertinggi yang dapat dilakukan secara sederhana dan cepat dalam memprediksi kelas dari data uji. Metode KNN memiliki konsep pengklasifikasian berdasarkan kedekatan jarak antar data sehingga mudah dipelajari dan pembentukan modelnya dapat dilakukan dengan cepat. Dari dua metode yang dipilih, metode Naïve Bayes menghasilkan akurasi 66,096%, sedangkan metode KNN menghasilkan akurasi 76,695%. Dengan membandingkan kedua metode, diperoleh kesimpulan bahwa metode KNN menghasilkan klasifikasi yang lebih baik. Berdasarkan hasil klasifikasi metode KNN terbaik, lima variabel yang paling berperan penting dalam pengklasifikasian adalah usia, status kehamilan, pendidikan tertinggi kepala rumah tangga, kepemilikan aset bergerak, dan kepemilikan aset tidak bergerak.

Kata kunci: kemiskinan, klasifikasi, PKH, Naïve Bayes, KNN

Abstract
Family Hope Program (PKH) is one the government's assistance programs as efforts to alleviate residents poverty which has been proven to be able to overcome chronic poverty in various countries. However, many people complain the determination of PKH recipients is not optimal yet, especially in Bantul Regency which is one of Yogyakarta's regency with the most poor residents. This research aimsto classify residents which have a right to be PKH recipients using Naïve Bayes and K-Nearest Neighbors (KNN) methods. Naïve Bayes method has a classifying concept based on the highest probability which can be done simply and fast in predicting the class of data testing. The KNN method has a classifying concept based on closeness distance between data so it is easy to learn and the model can be built quickly. From two of the chosen methods, Naïve Bayes method has accuration result of 66,09%, whereas KNN method accuration result of 76,695%. By comparing both methods, it can be concluded that the KNN method produces better classification results. Based on the best KNN classification results, the five variables which influence classification the most are age, pregnancy status, highest education head of household, movable assets ownership, and immovable assets ownership.

Keywords*: poverty, classification, PKH, Naïve Bayes, KNN*

PENDAHULUAN

Kemiskinan merupakan permasalahan global yang berasal dari berbagai sektor sehingga kompleks dan sulit untuk diselesaikan. Pemerintah Indoensia melakukan berbagai upaya pengentasan kemiskinan seperti dengan dibentuknya berbagai program bantuan untuk penduduk miskin. Salah satu program bantuan pemerintah, yaitu Program Keluarga Harapan (PKH) yang secara internasional disebut juga sebagai Conditional Cash Transfer (CCT), telah cukup berhasil menanggulangi kemiskinan kronis di berbagai negara (Kemensos, 2021). Namun, dalam realisasinya pemberian bantuan ini masih kurang tepat sasaran karena masih banyak penduduk miskin yang berhak menjadi penerima tidak terjaring ke dalam status penerima bantuan. Keluhan masyarakat ini salah satunya terjadi di Kabupaten Bantul, salah satu kabupaten di Provinsi Daerah Istimewa (DI) Yogyakarta dengan penduduk miskin terbanyak. Provinsi DI Yogyakarta sendiri selalu memiliki persentase kemiskinan lebih tinggi dari rata-rata persentase kemiskinan Indonesia sejak tahun 2012 (BPS, 2021). Atas dasar urgensi tersebut, perlu dilakukan penelitian lebih lanjut dalam bentuk penemuan klasifikasi yang memiliki akurasi tinggi dalam menentukan penerima PKH di Kabupaten Bantul, sehingga upaya penanggulangan kemiskinan dapat terlaksana dengan lebih optimal.

Bidang ilmu yang dapat menyelesaikan permasalahan klasifikasi data berjumlah besar dengan cepat dan tepat adalah data mining. Dalam data mining, terdapat beberapa metode yang dapat digunakan untuk mengklasifikasi, seperti Decision Trees, Aturan Induksi, Neural Network (NN), Naïve Bayes, dan K-Nearest Neighbors (KNN). Pada penelitian ini, metode yang digunakan adalah metode Naïve Bayes dan KNN. Pemilihan metode Naïve Bayes yang memiliki konsep pengklasifikasian dengan menghitung probabilitas tertinggi diperolehnya label dengan syarat atribut tertentu, didasarkan pada pertimbangan bahwa metode ini dapat mengklasifikasi dengan cepat dan sederhana. Kekurangan dari metode ini adalah data latih yang digunakan harus data lengkap sehingga dilakukan pembersihan data terlebih dahulu dengan cara menghapus data yang memiliki missing values. Metode KNN yang pengklasifikasiannya didasarkan pada kedekatan jarak antar data pada ruang berdimensi-n dipilih karena mudah untuk dipelajari dan modelnya dapat dibentuk dengan cepat. Walaupun memiliki kelemahan berupa diperlukannya penentuan nilai k yang tepat untuk mendapatkan hasil klasifikasi terbaik, hal ini dapat diatasi dengan klasifikasi berulang pada nilai k yang berbeda hingga ditemukan hasil klasifikasi dengan akurasi terbaik (Kotu & Deshpande, 2015).

Penelitian terkait yang membandingkan hasil klasifikasi kedua metode terpilih, yaitu metode Naïve Bayes dan metode KNN pernah dilakukan sebelumnya oleh Islam et al. (2007) untuk mengklasifikasi calon pendaftar yang berhak lolos menjadi pemilik kartu kredit. Penelitian tersebut menghasilkan kesimpulan bahwa metode KNN dengan nilai k = 5 memiliki akurasi sebesar 90,55%, lebih tinggi dari akurasi metode Naïve Bayes sebesar 87,57%. Namun, dengan akurasi yang tinggi, kedua metode ini baik untuk digunakan. Firasari et al. (2020) mengklasifikasi data penerima Bantuan Sosial (Bansos) di Desa Somokerto, Jawa Tengah, dengan menggunakan kedua metode dan menghasilkan kesimpulan bahwa metode KNN lebih baik dari metode Naïve Bayes dengan akurasi 89,04% berbanding dengan 87,67%.

Dari beberapa penelitian yang membandingkan akurasi hasil klasifikasi metode Naïve Bayes dan KNN, diketahui bahwa kedua metode memiliki hasil klasifikasi dengan akurasi yang hampir setara dan tinggi. Untuk memperkuat pernyataan tersebut, penelitian lain yang hanya menggunakan salah satu dari kedua metode pernah dilakukan sebelumnya. Nurmayanti et al. (2021) menggunakan metode Naïve Bayes untuk mengklasifikasi data penduduk miskin di Desa Lepak, Nusa Tenggara Barat dan memperoleh akurasi hasil klasifikasi sebesar 96,63%. Penelitian lain dengan metode yang sama oleh Purnama, Aziz, & Wiguna (2020) untuk mengklasifikasi penerima PKH di Desa Wae Jare tahun 2019 juga memperoleh akurasi tinggi yaitu sebesar 82,14%. Penelitian lain dengan menggunakan metode kedua, yaitu metode KNN,

dilakukan oleh Medjaded, Saadi, & Benyettou (2013) untuk mengklasifikasi diagnosa penyakit kanker. Penghitungan dilakukan dengan rumus jarak yang berbeda-beda dan memperoleh akurasi tertinggi dengan menggunakan rumus jarak Euclidean distance yaitu sebesar 98,70%. Kurnia et al. (2019) mengklasifikasi penduduk miskin dengan menggunakan metode yang sama, yaitu metode KNN, dan memperoleh akurasi sebesar 90% ketika k = 5, 7, dan 9.

Berdasarkan uraian permasalahan di atas, tujuan penelitian ini adalah untuk mengklasifikasi kelayakan penduduk miskin penerima PKH menggunakan metode Naïve Bayes dan KNN di Kabupaten Bantul. Pemilihan kedua metode didasarkan karena berdasarkan penelitian sebelumnya, akurasi yang dihasilkan tinggi dan layak untuk digunakan. Perbedaan penelitian ini dengan penelitian sebelumnya terletak pada objek penelitiannya yaitu di Kabupaten Bantul, serta penggunaan variabelnya, yaitu usia, kepemilikan disabilitas, kepemilikan penyakit kronis/menahun, status kehamilan, pendidikan tertinggi kepala rumah tangga, luas lantai bangunan, jenis lantai, jenis dinding, jenis atap, sumber air minum, sumber penerangan, fasilitas MCK (mandi, cuci, dan kakus), bahan bakar memasak, aset bergerak, aset tidak bergerak, kepemilikan hewan ternak, dan status kepesertaan PKH.

METODE

Deskripsi Data

Penelitian yang dilakukan merupakan penelitian kuantitatif pada data sekunder yang diperoleh dari instansi Badan Perencanaan dan Pembangunan Daerah (Bappeda) Kabupaten Bantul berupa Data Terpadu Kesejahteraan Sosial (DTKS) tahun 2020 yang berisi informasi mengenai 40% penduduk Kabupaten Bantul sebanyak 414.982 jiwa dengan status kesejahteraan sosial terendah.

Langkah Analisis Data

1. Pemilihan Data
   Pada tahap ini dilakukan pemilihan atribut yang digunakan pada penelitian. Atribut yang terpilih disesuaikan dengan kriteria penduduk miskin dengan beberapa tambahan atribut yang bepengaruh lainnya, yaitu usia, kepemilikan disabilitas, kepemilikan penyakit kronis/menahun, status kehamilan, pendidikan tertinggi kepala rumah tangga, luas lantai bangunan, jenis lantai, jenis dinding, jenis atap, sumber air minum, sumber penerangan, fasilitas MCK (mandi, cuci, dan kakus), bahan bakar memasak, aset bergerak, aset tidak bergerak, kepemilikan hewan ternak, dan status kepesertaan PKH.
2. Pra-pemrosesan Data
   Data yang memiliki *missing value* dihapus untuk memaksimalkan hasil kerja metode yang digunakan sehingga diperoleh tingkat akurasi penelitian yang maksimal.
3. Transformasi Data
   Data yang terpilih ditransformasi untuk menyederhanakan informasi sehingga klasifikasi dapat dilakukan dengan efisien dan cepat. Semua atribut kecuali atribut usia, aset bergerak, dan kepemilikan hewan ternak ditransformasi dengan pembobtan 1 yang menandakan sesuai dengan kriteria penerima PKH dan bobot 0 yang menandakan tidak sesuai kriteria. Atribut yang lain ditransformasi dengan menggunakan *Min-Max Normalization* sehingga data direpresentasikan dalam skala 0 hingga 1. Penghitungan normalisasi ini menurut Larose (2005) dengan $\bar{x}$ menandakan rata-rata sampel dan s sebagai standar deviasi sampel :

$$X^* = \frac{X - \bar{x}}{s}$$

4. Penambangan Data
   Data yang telah melalui proses pembersihan, pra-pemrosesan, dan transformasi kemudian diklasifikasi dengan metode *Naïve Bayes* dan KNN. Metode *Naïve Bayes* yang memiliki penghitungan berdasarkan probabilitas hubungan antara atribut dan label tertinggi memiliki rumus berikut (Han, Kamber, & Pei, 2012).

$$P(H|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|H)P(H)}{P(\boldsymbol{X})}, P(\boldsymbol{X}) > 0$$

   dengan definisi bahwa:
   $P(\boldsymbol{X})$: probabilitas terjadinya atribut $\boldsymbol{X}$
   $P(H)$: probabilitas terjadinya label H (probabilitas prior)
   $P(\boldsymbol{X}|H)$: probabilitas atribut $\boldsymbol{X}$ dengan syarat H (probabilitas kondisional)
   $P(H|\boldsymbol{X})$: probabilitas label H dengan syarat atribut $\boldsymbol{X}$ (probabilitas posterior)
   Metode KNN memiliki dasar klasifikasi berdasarkan kedekatan jarak antar data. Pada penelitian ini, digunakan penghitungan jarak dengan menggunakan rumus jarak Euclidean untuk suatu objek x ke y sebagai berikut (Santosa, 2007).

$$d(x,y) = \left|\left|x-y\right|\right|^2 = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

   Setelah klasifikasi dengan kedua metode, diterapkan tehnik *permutation importance*, yaitu tehnik pengukuran atribut penting berdasarkan tingkat kesalahan prediksi jika nilai suatu atribut yang digunakan diubah urutannya sehingga hubungan antara atribut dan label dipisahkan.
5. Interpretasi dan Evaluasi Data
   Hasil klasifikasi dievaluasi dengan menggunakan penghitungan persentase performa klasifikasi berupa persentase sensitivitas, spesifisitas, presisi, dan akurasi, serta pengecekan keakurasian klasifikasi dengan menggunakan *Cross Validation.* Hasil pengevaluasian kedua metode ini diinterpretasi dan dibandingkan sehingga dapat dibentuk kesimpulan metode mana yang lebih baik dari hasil penelitian secara keseluruhan.

HASIL DAN PEMBAHASAN

Hasil

Pengolahan data dilakukan dengan cara penambangan data klasifikasi dengan metode *Naïve Bayes* dan KNN. Pengolahan dilakukan dengan bahasa pemrograman Python. Setelah dilakukan klasifikasi, diterapkan *permutation importance* untuk mengetahui tingkat pengaruh atribut terhadap hasil klasifikasi. Selanjutnya, dengan mengevaluasi performa klasifikasi, hasil dari kedua metode dibandingkan dan dipilih metode dengan hasil terbaik.

1. Metode *Naïve Bayes*

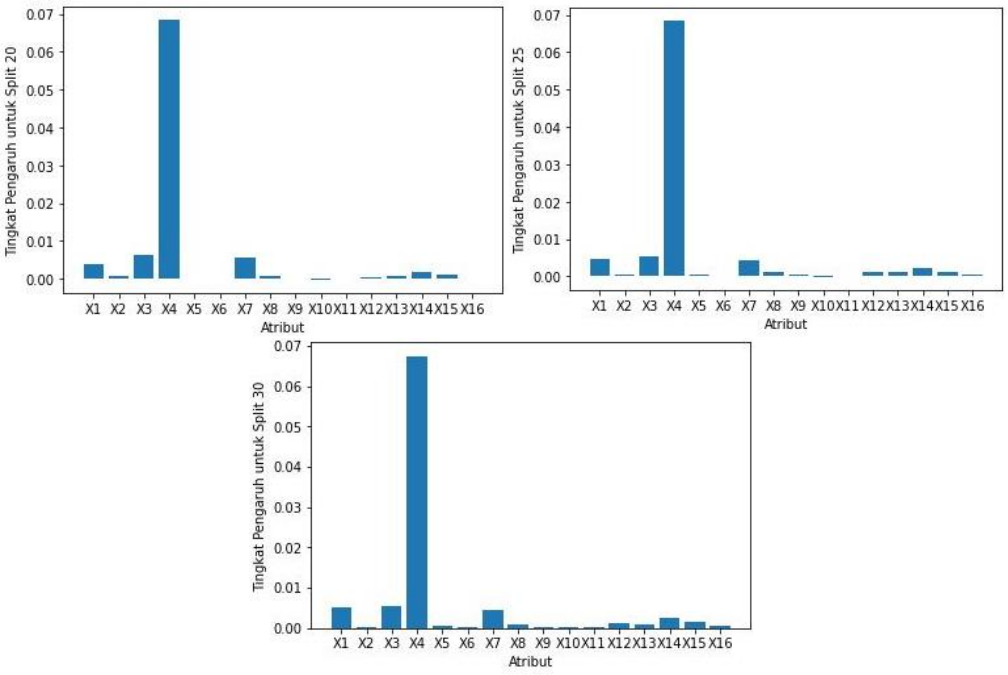
Klasifikasi pertama yang dilakukan adalah dengan menggunakan metode *Naïve Bayes.* Pengklasifikasian dilakukan pada tiga bentuk pembagian data latih dan data uji, yaitu pembagian 80:20, 75:25, dan 70:30. Metode yang digunakan adalah *Gaussian Naïve Bayes* karena terdapat atribut dalam bentuk kontinu. Selanjutnya, dilakukan pengecekan akurasi serta performa prediksi data latih dan data uji. Untuk dapat menghasilkan evaluasi kinerja yang lebih menyeluruh, dilakukan evaluasi tambahan menggunakan 10 *Fold-Cross Validation.* Langkah terakhir yang dilakukan pada tahap penambangan data ini adalah pengecekan atribut yang paling berpengaruh dalam pengklasifikan.

Tabel 1. Persentase Akurasi Hasil Penelitian Klasifikasi Metode Naïve Bayes

Keterangan	Akurasi Klasifikasi		
	20%	**25%**	**30%**
Data Latih	65,485%	65,488%	65,478%
Data Uji	66,096%	66,012%	65,961%

Berdasarkan Tabel 1, diketahui bahwa persentase akurasi terbaik klasifikasi data uji penduduk miskin penerima bantuan PKH dengan menggunakan metode *Naïve Bayes* sebesar 66,096%, yaitu ketika data uji yang digunakan sebanyak 20%. Akurasi pembentukan data latih pada pembagian data uji 20% ini adalah sebesar 65,485%.

Pada pengklasifikasian metode *Naïve Bayes* ini, dengan menggunakan tehnik *permutation importance*, diperoleh hasil bahwa lima variabel paling berpengaruh dalam pengklasifikasian adalah atribut usia (X1), kepemilikan penyakit kronis/menahun (X3), status kehamilan (X4), jenis lantai bangunan (X7), dan kepemilikan aset bergerak (X14). Gambar 1 menjelaskan histogram tingkat pengaruh sesuai atau tidaknya nilai dari masing-masing atribut dengan kriteria penduduk miskin penerima PKH terhadap pengklasifikasian. Semakin tinggi tingkat pengaruh atribut, maka semakin besar pengaruh atribut tersebut dalam pengklasifikasian penerima PKH.



Gambar 1. Histogram Tingkat Pengaruh Atribut terhadap Pengklasifikasian Naïve Bayes pada Masing-Masing Split

Dari hasil klasifikasi yang dilakukan, diperoleh Matriks Konfusi bagi data uji metode *Naïve Bayes* dengan perbandingan data latih dan data uji sebesar 80:20 yang menghasilkan akurasi tertinggi sebesar 66,096% adalah $\begin{bmatrix} 12246 & 1595 \\ 6481 & 3498 \end{bmatrix}$. Penghitungan evaluasi persentase performa klasifikasi yaitu sebagai berikut.

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% = \frac{12246}{12246 + 1595} \times 100\% = 88{,}476\%$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% = \frac{3498}{3498 + 6481} \times 100\% = 35{,}054\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\% = \frac{12246}{12246 + 6481} \times 100\% = 65{,}392\%$$

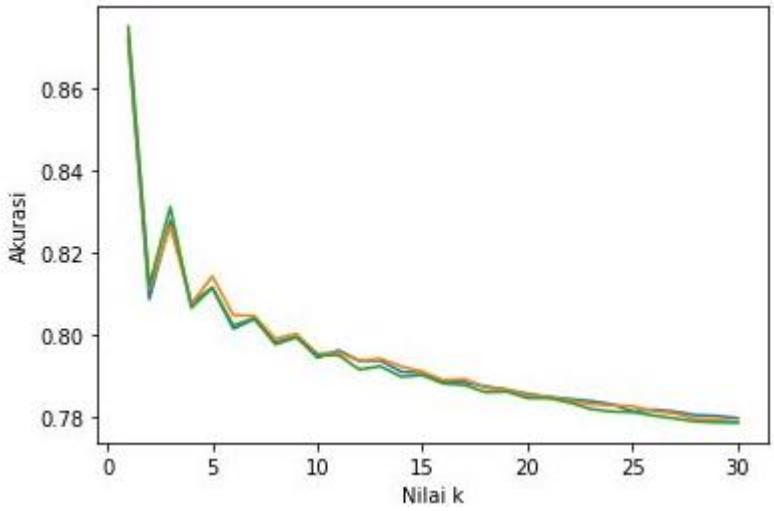
Arti dari hasil penghitungan ini menunjukkan bahwa metode *Naïve Bayes* memiliki akurasi prediksi label sebesar 66,096% dengan kemampuannya untuk memilih data yang sesuai sebesar 88,476%, kemampuan untuk menolak data yang tidak sesuai sebesar 35,054%, dan proporsi banyaknya kasus yang diprediksi benar sebesar 65,392%.

Klasifikasi metode *Naïve Bayes* dengan pembagian data 80:20 menghasilkan rata-rata persentase *Cross Validation* data latih sebesar 65,931% dan data uji sebesar 65,453% yang memiliki arti bahwa metode *Naïve Bayes* dapat membentuk data latih dengan akurasi 65,931% dan keakurasian hasil klasifikasi data uji sebesar 65,453%. Persentase ini tidak berbeda jauh dengan persentase akurasi pengklasifikasian data latih sebesar 65,485% dan data uji 66,096% sehingga dapat dikatakan bahwa *Cross Validation* tidak meningkatkan performa metode karena akurasi yang diperoleh adalah akurasi maksimum metode ini dalam mengklasifikasi data.

2. Metode KNN

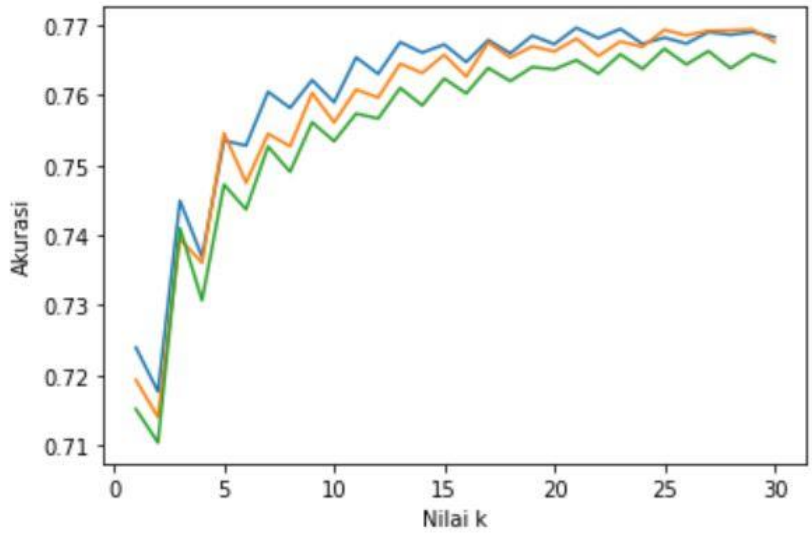
Klasifikasi dilakukan dengan pembagian data latih dan data uji menjadi tiga perbandingan seperti pada tahapan metode *Naïve Bayes* sebelumnya, yaitu 80:20, 75:25, dan 70:30. Nilai *k* yang digunakan adalah bilangan dari 1 hingga 30. Pemilihan nilai *k* dari 1 hingga 30 ini dikarenakan setelah melakukan percobaan klasifikasi berulang dengan menambahkan nilai *k* satu per satu dimulai dari *k* = 1, dapat terlihat pola bahwa akurasi klasifikasi menurun ketika nilai *k* > 30. Agar dapat memperoleh jarak terdekat antar objek, maka penghitungan jarak dilakukan dengan menggunakan penghitungan jarak Euclidean. Hasil klasifikasi dari metode tersebut dihitung akurasinya. Metode dengan nilai *k* yang menghasilkan akurasi prediksi data uji tertinggi selanjutnya dicek performa serta kinerjanya dengan menggunakan Matriks Konfusi dan 10 *Fold-Cross Validation.* Selanjutnya, langkah terakhir yang dilakukan pada tahap ini adalah pengecekan atribut yang paling berpengaruh dalam pengklasifikan.

Gambar 2 menunjukkan grafik persentase akurasi pembagian data latih masing-masing perbandingan klasifikasi KNN dengan nilai *k* bervariasi dari 1 hingga 30.



Gambar 2. Grafik Persentase Akurasi Data Latih Hasil Klasifikasi Metode KNN

Gambar 3 menunjukkan grafik persentase akurasi pembagian data uji masing-masing perbandingan klasifikasi KNN dengan nilai *k* bervariasi dari 1 hingga 30.

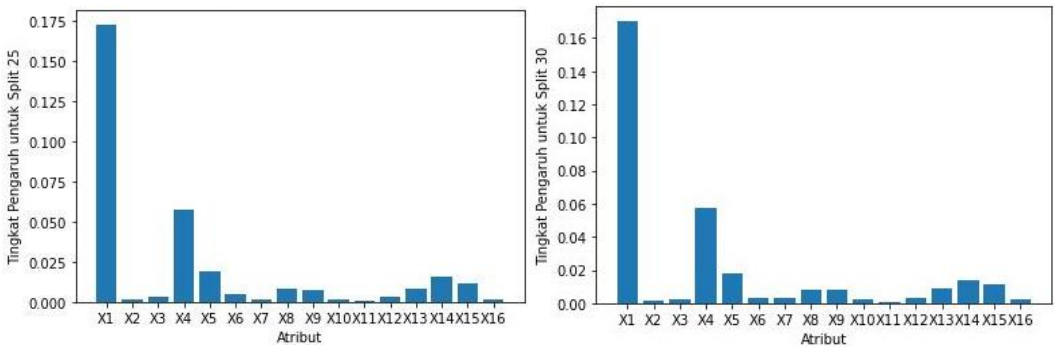


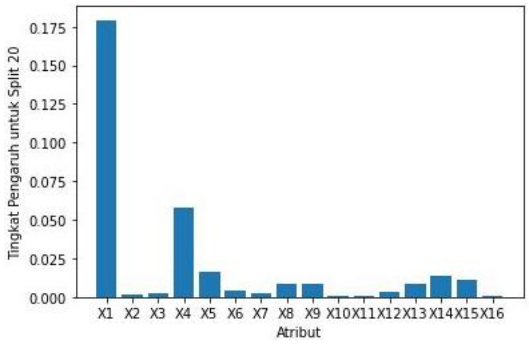
Gambar 3. Grafik Persentase Akurasi Data Uji Hasil Klasifikasi Metode KNN

Perbedaan warna garis pada Gambar 2 dan Gambar 3 menandakan perbedaan perbandingan pembagian data latih dan data uji dalam pengklasifikasian menggunakan metode KNN: garis biru menandakan perbandingan 80:20, garis kuning menandakan perbandingan 75:25, dan garis hijau menandakan perbandingan 70:30.

Dari penelitian klasifikasi penduduk miskin penerima bantuan PKH dengan menggunakan metode KNN yang telah dilakukan, terlihat pola bahwa akurasi maksimal diperoleh ketika nilai *k* ganjil dengan persentase akurasi terbaik adalah 76,965% ketika nilai *k* = 21 dengan data uji yang digunakan sebanyak 20%. Akurasi pembentukan data latihnya adalah 78,503%.

Pada pengklasifikasian metode KNN ini, dengan menggunakan tehnik *permutation importance,* diperoleh hasil bahwa lima variabel paling berpengaruh dalam pengklasifikasian adalah atribut usia (X1), status kehamilan (X4), pendidikan tertinggi kepala rumah tangga (X5), kepemilikan aset bergerak (X14), dan kepemilikan aset tidak bergerak (X15). Gambar 4 menjelaskan histogram tingkat pengaruh sesuai atau tidaknya nilai dari masing-masing atribut dengan kriteria penduduk miskin penerima PKH terhadap pengklasifikasian. Semakin tinggi tingkat pengaruh atribut, maka semakin besar pengaruh atribut tersebut dalam pengklasifikasian penerima PKH.





Gambar 4. Histogram Tingkat Pengaruh Atribut terhadap Pengklasifikasian KNN pada Masing-Masing Split

Matriks konfusi dari klasifikasi terbaik metode KNN dengan perbandingan data 80:20 dan nilai k = 21 yang menghasilkan akurasi sebesar 76,965% adalah $\begin{bmatrix} 10558 & 3045 \\ 2442 & 7745 \end{bmatrix}$. Penghitungan evaluasi persentase performa klasifikasi yaitu sebagai berikut.

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% = \frac{10588}{10588 + 3045} \times 100\% = 77{,}664\%$$

$$Specificity = \frac{TN}{TN + FP} \times 100\% = \frac{7745}{7745 + 2442} \times 100\% = 76{,}028\%$$

$$Precision = \frac{TP}{TP + FP} \times 100\% = \frac{10588}{10588 + 2442} \times 100\% = 81{,}215\%$$

Artinya, metode KNN memiliki akurasi prediksi label sebesar 76,965% dengan kemampuannya untuk memilih data yang sesuai sebesar 77,664%, kemampuan untuk menolak data yang tidak sesuai sebesar 76,028%, dan proporsi banyaknya kasus yang diprediksi benar sebesar 81,215%.

Klasifikasi metode KNN dengan pembagian data 80:20 menghasilkan rata-rata persentase *Cross Validation* baik data latih maupun data uji sebesar 76,073%, yang artinya bahwa metode KNN dapat membentuk data latih dan data uji dengan akurasi 76,073%. Persentase ini tidak berbeda jauh dengan persentase akurasi pengklasifikasian data latih sebesar 78,503% dan data uji 76,965%. Artinya, *Cross Validation* tidak meningkatkan performa metode sehingga akurasi tersebut adalah akurasi maksimum metode ini dalam mengklasifikasi data.

Pembahasan

Berdasarkan penelitian yang telah dilakukan, diperoleh hasil bahwa pengklasifikasian metode KNN dapat mengklasifikasi penduduk miskin penerima PKH di Kabupaten Bantul lebih baik dengan besar persentase keakurasian sebesar 76,965% jika dibandingkan dengan metode *Naïve Bayes* yang menghasilkan akurasi sebesar 66,096%. Hasil ini sesuai dengan hasil penelitian Sihombing & Arsani (2021) yang menggunakan perbandingan empat metode untuk mengklasifikasi penduduk miskin di Indonesia pada tahun 2018: *Decision Tree, Naïve Bayes,* KNN, dan *Rotation Forest.* Pada penelitian tersebut akurasi tertinggi juga diperoleh jika menggunakan metode KNN baik berdasarkan persentase akurasinya, maupun persentase performa metodenya. Penelitian oleh Islam et al. (2007) yang melakukan perbandingan hasil klasifikasi metode *Naïve Bayes* dan KNN pada data calon pemilik kartu kredit juga menghasilkan kesimpulan bahwa klasifikasi dengan metode KNN memiliki akurasi lebih tinggi dibandingkan dengan akurasi metode *Naïve Bayes,* yaitu 90,55% dengan akurasi metode *Naïve Bayes* sebesar 87,57%. Selain itu, lebih tingginya akurasi klasifikasi metode KNN

dibandingkan metode *Naïve Bayes* juga sesuai dengan penelitian Laksana (2020) yang mengklasifikasi komentar positif dan negatif di Twitter tentang Dewan Perwakilan Rakyat (DPR) dengan akurasi metode KNN 80% sedangkan akurasi metode *Naïve Bayes* 77%.

Tingginya akurasi metode KNN ini diperoleh karena klasifikasi dengan metode KNN dilakukan berulang kali dengan nilai *k* yang berbeda hingga diperoleh akurasi tertinggi. Cara ini sesuai dengan penelitian Moldagulova & Sulaiman (2017) yang menerapkan metode KNN untuk klasifikasi dokumen dalam bentuk teks dengan nilai *k* yang bervariasi. Pada penelitian tersebut, diperoleh hasil bahwa nilai *k* yang berbeda dapat menghasilkan akurasi yang berbeda juga sehingga penting untuk mengklasifikasi dengan nilai *k* bervariasi hingga menghasilkan akurasi tertinggi. Pengklasifikasian menggunakan metode KNN yang dilakukan secara berulang dengan nilai *k* yang berbeda juga sesuai dengan penelitian yang dilakukan oleh Jabbar, Deekshatulu, & Chandra (2013) untuk mengklasifikasi data penyakit jantung.

SIMPULAN

Kesimpulan yang dapat diambil atas hasil klasifikasi kelayakan penduduk miskin penerima Program Keluarga Harapan (PKH) di Kabupaten Bantul adalah sebagai berikut.

1. Klasifikasi kelayakan penduduk miskin penerima bantuan PKH dengan menggunakan metode *Naïve Bayes* dan *K-Nearest Neighbors* (KNN) dilakukan dengan melakukan pemilihan atribut yang sesuai dengan kriteria penduduk miskin, yaitu usia, kepemilikan disabilitas, kepemilikan penyakit kronis, status kehamilan, pendidikan tertinggi kepala rumah tangga, luas lantai, jenis lantai, jenis dinding, jenis atap, sumber air minum, sumber penerangan, fasilitas MCK, bahan bakar memasak, aset bergerak, aset tidak bergerak, dan banyak ternak. Data yang digunakan adalah data sekunder Data Terpadu Kesejahteraan Sosial (DTKS) tahun 2020 yang berisi 40% penduduk Kabupaten Bantul dengan status kesejahteraan sosial terendah, yaitu sebanyak 414.982 penduduk. Kumpulan atribut terpilih dari data tersebut yang memiliki *missing values* dihapus agar pengklasifikasian lebih optimal karena data yang digunakan adalah data yang lengkap. Selanjutnya, tahap transformasi data dengan melakukan pembobotan untuk data kategorik berupa bobot 1 menyatakan memenuhi sebagai penerima PKH dan bobot 0 menyatakan tidak memenuhi sebagai penerima PKH. Beberapa atribut lainnya yang ada dalam bentuk data numerik diubah skalanya dengan menerapkan normalisasi data sehingga ada dalam skala 0 hingga 1. Setelah tahap ini, data diolah dengan menggunakan metode *Naïve Bayes* dan KNN. Kemudian, hasil klasifikasi diinterpretasi dan dievaluasi performanya.
2. Metode *Naïve Bayes* menghasilkan akurasi tertinggi untuk data tes sebesar 66,096% pada pembagian data 80:20 dengan akurasi data latihnya sebesar 65,485%. Metode KNN menghasilkan akurasi tertinggi untuk data tes sebesar 76,965% dengan akurasi data latihnya sebesar 78,503%. Jadi, dapat disimpulkan bahwa metode KNN lebih baik jika dibandingkan dengan metode *Naïve Bayes* untuk mengklasifikasi kelayakan penduduk miskin penerima PKH di Kabupaten Bantul. Performa hasil klasifikasi yang dilakukan oleh metode KNN memiliki sensitivitas 77,664%, spesifisitas 76, 028%, dan presisi 81,215%.

Saran yang dapat diberikan untuk penelitian selanjutnya adalah dengan dilakukannya penerapan metode klasifikasi *data mining* lain seperti *Decision Tree, Neural Network* (NN), Aturan Induksi, maupun kombinasi dari metode klasifikasi yang ada sebagai perbandingan sehingga dapat diketahui metode mana yang menghasilkan klasifikasi terbaik.

UCAPAN TERIMA KASIH

Terimakasih kepada koordinator Prodi Maatematika dan seluruh Dosen Prodi Matematika yang telah memberikan ilmu dan bimbingan hingga terselesainya artikel ini.

DAFTAR PUSTAKA

Badan Pusat Statistik. (2021). Persentase Penduduk Miskin (P0) Menurut Provinsi dan Daerah (2007-2021). Diakses dari https://www.bps.go.id/indicator/23/192/1/persentase-penduduk-miskin-p0-menurut-provinsi-dan-daerah.html.

Firasari, E. et al. (2020). Comparation of K-Nearest Neighbor (KNN) and Naïve Bayes Algorithm for the Classification of the Poor in Recipients of Social Assistance. Jurnal of Physics: Conference Series, 1641, 1-6. doi: 10.1088/1742-6596/1641/1/012077.

Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques (Third Edition). Massachusetts: Morgan Kauffman Publishers.

Islam, M.J. et al. (2007). Investigating the Performance of Naïve-Bayes Classifiers and K-Nearest Neighbors Classifiers. International Conference on Convergence Information Technology, 4, 1541-1546. doi: 10.1109/ICCIT.2007.148.

Jabbar, M.A., Deekshatulu, B.L., & Chandra, P. (2013). Classification of Heart Disease Using K-Nearest Neighbor and Genetic Algorithm. International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA) 2013, 10, 85-94. doi: 10.1016/j.protcy.2013.12.340.

Kementerian Sosial Republik Indonesia. (2021). Program Keluarga Harapan (PKH). Diakses dari https://kemensos.go.id/program-keluarga-harapan-pkh.

Kotu, V. & Deshpande, B. (2015). Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner. Massachusetts: Esevier Inc.

Kurnia, F. et al. (2019). Klasifikasi Keluarga Miskin Menggunakan Metode K-Nearest Neighbors Berbasis Euclidean Distance. Seminar Nasional Teknologi, Informasi, Komunikasi, dan Industri (SNTIKI), 11, 230-239. Diakses dari http://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/download/8089/4475.

Laksana, T.G. et al. (2020). Classification of Twitter Comments About the Image of the People's Representative Council (DPR) Using the K-Nearest Neighbor (KNN) Method and Naïve Bayes. International Conference of Global Education and Society Science (ICOGESS) 2019. doi: 10.4108/eai.14-3-2019.2292042.

Larose, D.T. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. New Jersey: John Wiley & Sons Inc.

Medjahed, S.A., Bote, M.P., & Deshmukh, S.D. (2013). Heart Disease Prediction System Using Naïve Bayes. International Journal of Enhanced Research in Science Technology & Engineering, 2, 1-5. doi: 10.1.1.378.9860.

Moldagulova, A. & Sulaiman, R.B. (2017). Using KNN Algorithm for Classification of Textual Documents. 8th International Conference on Information Technology (ICIT), 665-671. doi: 10.1109/ICITECH.2017.8079924.

Nurmayanti, W.P. et al. (2021). Penerapan Naïve Bayes dalam Mengklasifikasikan Masyarakat Miskin di Desa Lepak. Jurnal Kajian Ilmu dan Pendidikan Geografi, 5, 123-132. doi: 10.29408/geodika.v5i1.3430.

Purnama, A.I., Aziz, A., & Wiguna, A.S. (2020). Penerapan Data Mining untuk Mengklasifikasi Penerima Bantuan PKH Desa Wae Jare Menggunakan Metode Naïve Bayes. KURAWAL Jurnal Teknologi, Informasi dan Industri, 3, 173-180. doi: 10.33479/KURAWAL.2020.3.2.173.

Santosa, B. (2007). Data Mining Teori dan Aplikasi. Yogyakarta: Graha Ilmu.

Sihombing , P.R. & Arsani, A.M. (2021). Comparison of Machine Learning Methods in Classifying Poverty in Indonesia in 2018. Jurnal Teknik Informatika (JUTIF), 2, 51-56. doi: 10.20884/1.jutif.2021.2.1.52.