

Perbandingan Algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dalam Menentukan Kriteria Masyarakat Miskin

Arif Santoso¹⁾, Ema Utami²⁾, Anggit Dwi Hartanto³⁾

Magister Teknik Informatika, Universitas AMIKOM Yogyakarta,
Jalan Ring Road Utara, Condong Catur, Depok, Sleman, Yogyakarta 55281
¹arif.1261@students.amikom.ac.id, ²ema.u@amikom.ac.id, ³anggit@amikom.ac.id

Abstrak

Mewabahnya Virus Covid-19 memberikan dampak yang sangat besar dibidang ekonomi. Melemahnya ekonomi masyarakat menjadi permasalahan serius yang sangat perlu segera diatasi. Dalam upaya pemulihan ekonomi, pemerintah mengeluarkan kebijakan-kebijakan untuk pengentasan ekonomi antara lain penyaluran BLT. Akan tetapi kebijakan tersebut justru menimbulkan masalah baru yaitu penyaluran yang tidak tepat sasaran. Hal ini menimbulkan gejala dimasyarakat. Penelitian ini bertujuan untuk mengklasifikasikan data sesuai dengan kriteria dan memperoleh hasil terbaik dua metode yang akan digunakan. Penelitian ini menggunakan dua algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dengan data yang diperoleh dari Data terpadu kesejahteraan Sosial (DTKS) dengan dua variabel yaitu mampu dan miskin.

Hasil klasifikasi dengan dua algoritma *Naïve Bayes* dan *K-Nearest Neighbor* diperoleh hasil masing-masing 72.64% dan 95.40% dengan nilai AUC 0.836 dan 0.877. Berdasarkan nilai AUC yang diperoleh kedua algoritma tingkat akurasi termasuk *good classification*. Algoritma *K-Nearest Neighbor* lebih baik dalam klasifikasi masyarakat miskin dibandingkan algoritma *Naïve Bayes* dengan akurasi 95.40% dan nilai AUC 0.877.

Kata kunci: Klasifikasi, *Naïve Bayes*, k-NN, Kemiskinan, Perbandingan

Abstract

The outbreak of the Covid-19 virus has had a huge impact on the economy. The weakening of the community's economy is a serious problem that urgently needs to be addressed. In an effort to recover the economy, the government issued policies for economic alleviation, including the distribution of BLT. However, this policy actually creates a new problem, namely distribution that is not right on target. This caused turmoil in society. This study aims to classify the data according to the criteria and obtain the best results from the two methods that will be used. This study uses two *Naïve Bayes* algorithms and *K-Nearest Neighbor* with data obtained from the Integrated Social Welfare Data (DTKS) with two variables, namely the rich and the poor.

The results of the classification using the two algorithms *Naïve Bayes* and *K-Nearest Neighbor* obtained the results respectively 72.64% and 95.40% with AUC values of 0.836 and 0.877. Based on the AUC value obtained, the two algorithms have a good classification level of accuracy. The *K-Nearest Neighbor* algorithm is better in classifying the poor than the *Naïve Bayes* algorithm with an accuracy of 95.40% and an AUC value of 0.877.

Keywords: Classification, *Naïve Bayes*, k-NN, Poverty, Comparison

1. PENDAHULUAN

Mewabahnya Virus Covid-19 yang muncul pada akhir tahun 2019 hingga saat ini memberikan dampak yang sangat besar. Hampir semua sektor kehidupan merasakan pengaruh yang sangat signifikan. Sektor ekonomi menjadi salah satu sektor yang paling terdampak akibat mewabahnya Virus

covid-19 ini. Melemahnya ekonomi masyarakat menjadi permasalahan serius yang sangat perlu segera diatasi. Dalam upaya pemulihan ekonomi, pemerintah mengeluarkan kebijakan-kebijakan untuk pengentasan ekonomi. Kebijakan-kebijakan yang diluncurkan pemerintah antara lain penyaluran BLT atau bantuan langsung tunai.

Program ini berupa penyaluran bantuan secara tunai yang dimaksudkan untuk dibelanjakan kebutuhan pokok. Akan tetapi kebijakan tersebut justru menimbulkan masalah baru yaitu penyaluran yang tidak tepat sasaran. Hal ini menimbulkan gejala dimasyarakat. Dari permasalahan yang muncul perlu dilakukan klasifikasi data dengan kriteria-kriteria yang dimiliki. Hal ini bertujuan untuk memisahkan dan mengelompokan masyarakat berdasarkan komponen yang dimiliki, sehingga upaya pemerintah melakukan pengentasan kemiskinan dapat berjalan tepat sasaran. Proses klasifikasi yang dilakukan dengan menggunakan algoritma Naïve Bayes dan KNN. Pada tahap klasifikasi ini akan dilakukan perbandingan tingkat akurasi yang dihasilkan.

2. METODE PENELITIAN

Jenis penelitian yang akan dilakukan adalah eksperimen dengan tujuan mengetahui tingkat akurasi yang dihasilkan oleh algoritma Naive bayes dan K-Nearest Neighbor. Sifat dari penelitian ini yaitu bersifat *deskriptif*. Penelitian ini menjelaskan tahapan-tahapan klasifikasi dari metode yang digunakan sampai dengan diperoleh hasil klasifikasi. Pendekatan penelitian ini merupakan pendekatan kuantitatif. Pemilihan pendekatan ini dalam penelitian karena penelitian yang akan dilakukan menghasilkan data berbentuk angka. Angka-angka yang dihasilkan merupakan gambaran hasil proses dari penerapan metode yang digunakan dalam penelitian. Penelitian yang akan dilakukan dibedakan menjadi dua tahapan antara lain:

a. Pengumpulan Data

Pada tahap ini penulis melakukan studi literatur tentang informasi yang dibutuhkan. Informasi dan data yang digunakan sebagai acuan dalam penelitian diambil dari buku, jurnal dan internet.

b. Analisis Data

Setelah data diperoleh kemudian dilanjutkan dengan analisis data sebagai berikut:

1) *Prepare Data*

Merupakan tahap awal dalam Analisis data, Peneliti menyiapkan data yang akan digunakan dalam penelitian.

2) *Cleansing Data*

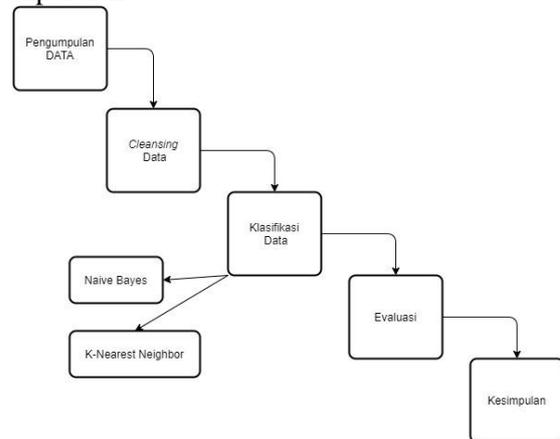
Setelah data didapatkan kemudian dilakukan *Cleansing data*. Proses ini dilakukan untuk memisahkan data yang tidak valid yang dapat mempengaruhi hasil yang dihasilkan dalam penelitian.

3) Klasifikasi

Proses klasifikasi dilakukan menggunakan dua algoritma yaitu *Naive Bayes* dan *K-Nearest Neighbour*. Penggunaan dua metode ini bertujuan untuk memperoleh hasil akurasi klasifikasi terbaik.

4) Evaluasi

Setelah tahapan demi tahapan dilewati kemudian proses terakhir adalah evaluasi. Proses ini menghasilkan kesimpulan yang dapat ditarik dari proses klasifikasi yang diperoleh.



Gambar 1. Alur Penelitian

Pada gambar 1 diatas merupakan gambaran tentang alur penelitian mulai dari awal pengumpulan data, pemrosesan data samapai dengan memperoleh hasil penelitian dari metode-metode yang digunakan.

3. TINJAUAN PUSTAKA

Beberapa tahun sebelumnya telah dilakukan beberapa penelitian yang berkaitan dengan klasifikasi kemiskinan. Penelitian pertama tentang komparasi metode naive bayes dan K-NN dalam klasifikasi masyarakat miskin. Penelitian ini bertujuan untuk memperoleh hasil paling akurat dari dua metode yang digunakan sekaligus menggantikan klasifikasi manual yang selama ini dirasa kurang akurat. Dari proses klasifikasi diperoleh hasil akurasi naive bayes sebesar 89,04% dan K-NN sebesar 87,67%.

Dengan demikian dapat disimpulkan Algoritma Naive Bayes lebih baik dalam klasifikasi masyarakat miskin. (Firasari & dkk, 2020)

Penelitian pertama ini memiliki kemiripan dengan penelitian yang akan dilakukan. Akan tetapi penelitian ini tidak memberikan kejelasan mengenai kriteria yang digunakan dalam menentukan status kemiskinan.

Penelitian kedua tentang pertumbuhan ekonomi yang tidak merata. Penelitian ini melakukan klasifikasi berdasarkan data penduduk berdasarkan umur, Pendidikan, Pekerjaan, Penghasilan, Tanggungan, Status. Metode yang digunakan adalah metode *Naive Bayes Classifier* dengan akurasi 73%, dengan *Precision* 92% dan *Recall* 86%. (Annur, 2018)

Penelitian ketiga tentang penyaluran bantuan yang tidak sesuai sasaran. Penelitian ini bertujuan untuk menciptakan aplikasi yang dapat mengklasifikasi keluarga miskin dengan algoritma KNN. Prinsip kerja algoritma KNN adalah mencari jarak terdekat antara data yang diuji dan data latihnya. Dari 100 contoh data yang digunakan akurasi yang diperoleh pada nilai $K=5$, $K=7$ dan $K=9$ dengan perbandingan 90:10 diperoleh akurasi sebesar 90%. (Kurna, Jhoni, Ichsan, & Siti, 2019)

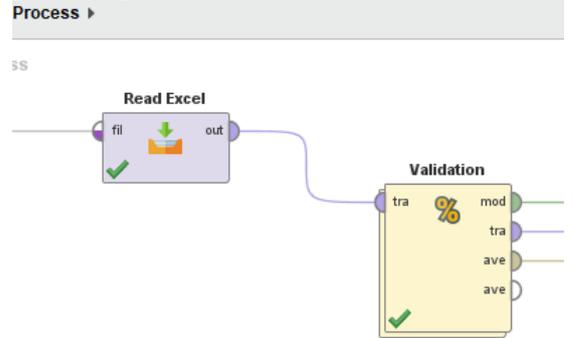
Penelitian keempat tentang kriteria kemiskinan di daerah Tabanan yang tidak tepat sasaran. Peneliti membangun aplikasi yang dapat mengidentifikasi rumah tangga yang berada di kabupaten tersebut. Sistem yang dibangun menerapkan algoritma KNN dan menghasilkan akurasi sebesar 83%. (Supriana & Luh, 2019)

Dari penelitian di atas memiliki kemiripan dari objek dan algoritma yang digunakan dalam penelitian yang akan dilakukan. Perbedaan yang dimiliki yaitu penelitian yang akan dilakukan ini membandingkan dua algoritma klasifikasi yaitu Naive Bayes dan KNN untuk memperoleh hasil yang terbaik.

4. HASIL DAN PEMBAHASAN

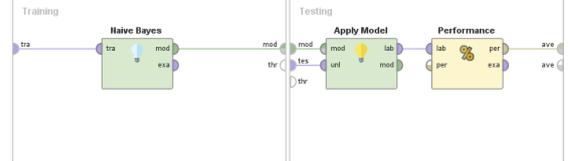
Dataset yang digunakan dalam penelitian diperoleh dari data terpadu kesejahteraan sosial (DTKS). Dataset yang akan digunakan terdiri dari 1378 data dengan 10 atribut. Dataset kemudian diolah

menggunakan *tools* RapidMiner dengan desain proses seperti pada gambar 2 dibawah ini.

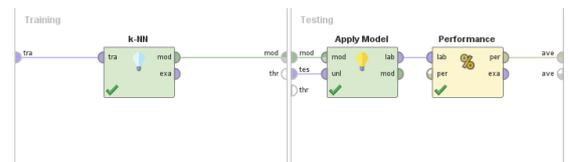


Gambar 2. Desain Proses

Read excel merupakan fitur untuk membaca file berbentuk excel. Operator split validation digunakan untuk membagi data menjadi dua bagian yaitu data *training* dan data *testing*. Parameter pada operator split validation dengan *number of validations* 0.7 dan *sampling type stratified sampling*. *Stratified sampling* membagi subset secara acak dengan memperhatikan distribusi kelas sama dengan dataset keseluruhan. Proses klasifikasi menggunakan dua algoritma yaitu *Naive Bayes* dan K-Nearest Neighbor seperti pada gambar 3 dan gambar 4.



Gambar 3. Klasifikasi Naive Bayes



Gambar 4. Klasifikasi K-Nearest Neighbor

Pada gambar 3 dan gambar 4 diatas dapat kita lihat pembagian data menjadi dua bagian yaitu *training* dan *testing*. Operator Naive Bayes dan K-NN digunakan untuk memodelkan proses klasifikasi. Pada penelitian ini penulis menggunakan nilai K untuk algoritma K-Nearest Neighbor =10. Operator apply model bertugas menterjemahkan pemodelan yang telah dilakukan oleh Naive Bayes dan K-NN pada bagian training untuk kemudian dapat diterapkan pada data *testing*. Performance dalam merupakan operator yang berfungsi

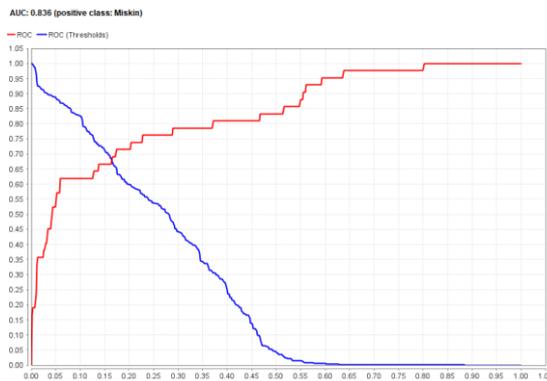
untuk menampilkan secara visual apa yang telah dimodelkan dalam bentuk vektor.

Setelah proses selesai maka diperoleh hasil Klasifikasi algoritma *Naïve Bayes* memperoleh hasil akurasi 72.64% dan nilai AUC sebesar 0.836 ditunjukkan pada gambar 5 dan gambar 6 di bawah ini.

accuracy: 72.64%

	true Mampu	true Mistin	class precision
pred. Mampu	288	10	96.40%
pred. Mistin	103	32	23.70%
class recall	72.24%	76.19%	

Gambar 5. Hasil Klasifikasi *Naïve Bayes*



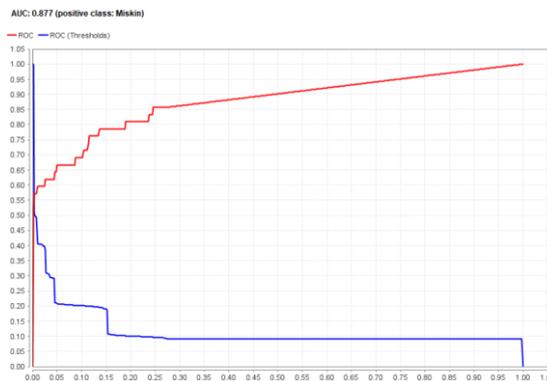
Gambar 6. Grafik AUC *Naïve Bayes*

Kemudian Algoritma K-nearest neighbor memperoleh hasil akurasi 95.40% dengan nilai AUC 0.877 seperti gambar 7 dan 8 di bawah ini:

accuracy: 95.40%

	true Mampu	true Mistin	class precision
pred. Mampu	370	18	95.36%
pred. Mistin	1	24	96.00%
class recall	99.73%	57.14%	

Gambar 7. Hasil Klasifikasi K-Nearest Neighbor



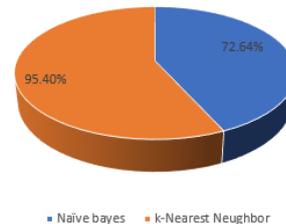
Gambar 6. Grafik AUC K-Nearest Neighbor

Hasil yang diperoleh pada penelitian dengan dua algoritma ini dapat kita lihat pada tabel 1 di bawah ini:

Tabel 1. Hasil Klasifikasi

	Naive Bayes	K-Nearest Neighbor
Akurasi	72.64%	95.40%
AUC	0.836	0.877

Berdasarkan nilai AUC yang diperoleh pada tabel 1 di atas kedua algoritma masuk kedalam kategori *good clasification*. Perbandingan Algoritma *Naïve Bayes* dan K-Nearest Neighbor menunjukkan algoritma K-Nearest Neighbor lebih baik dalam mengklasifikasikan kategori masyarakat miskin. Perbandingan kedua algoritma dapat dilihat pada gambar 8 di bawah ini:



Gambar 8. Perbandingan Hasil

5. PENUTUP

Dari tahapan penelitian yang telah dilakukan dengan dua algoritma *Naïve Bayes* dan K-Nearest Neighbor diperoleh hasil masing-masing 72.64% dan 95.40% dengan nilai AUC 0.836 dan 0.877. Berdasarkan nilai AUC yang diperoleh kedua algoritma tingkat akurasi termasuk *good classification*.

Berdasarkan hasil akurasi dan nilai AUC yang dihasilkan dalam penelitian algoritma K-Nearest Neighbor lebih baik dalam klasifikasi masyarakat miskin dibandingkan algoritma *Naïve Bayes* dengan akurasi 95.40% dan nilai AUC 0.877.

6. REFERENSI

Annur, H. (2018). Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes. *ILKOM Jurnal Volume 10 Nomor 2*.

- Antika, P. D., & Wowon, P. (2020). *Modul data Mining*. Jakarta: Fakultas Ilmu Komputer, Universitas Bhayangkara.
- Denis, A. C., Donny, A. B., Lia, A., & I Wayan, S. W. (2013). *Belajar Data Mining dengan RapidMiner*. Jakarat: Remi Sanjaya.
- Firasari, E., & dkk. (2020). Comparation of K-Nearest Neighbor (K-NN) and Naive Bayes Algorithm for the Classification of the poor in Recipients of Social Assistance. *Journal of Physics: Confereace Series ICASD*.
- Ismail, A. M. (2018, August 17). Cara Kerja Algoritma k-Nearest Neighbor (k-NN). <https://www.reserchgate.net/publication/330840826>.
- Kurna, F., Jhoni, K., Ichsan, F., & Siti, M. (2019). Klasifikasi Keluarga Miskin Menggunakan Metode K-Nearest Neighbor Berbasis euclidian Distance. *seminar nasional teknologi informasi, komunikasi dan industri (SNTIKI) 11, Pekanbaru*.
- Pemerintah Provinsi Jawa Tengah. (2019). *Satu Perangkat Daerah Satu Desa Dampingan Menuju Desa Lebih Sejahtera*. Semarang.
- Purnomo, I. I. (2016). Klasifikasi Status Kesejahteraan Rumah Tangga Menggunakan Algoritma K-Nearest Neighbor dan Seleksi Fiture Berbasis Chi Squard. *Technologia, ol 7, No3*.
- Supriana, I. W., & Luh, G. A. (2019). Implementasi K-Nearest Pada Penentuan Keluarga Miskin Bagi Dinas Sosial Kabupaten Tabanan. *Jurnal Teknologi Informasi dan Komputer, Volume 5, Nomor 1*.