

Application of the Naïve Bayes Algorithm in Predicting Acceptance of Family Hope Program Assistance (PKH)

Penerapan Algoritma Naïve Bayes dalam Prediksi Penerimaan Bantuan Program Keluarga Harapan (PKH)

Allan Wahyu Sekti Sriwibowo

Prodi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Ponorogo

Aslan Alwi

Prodi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Ponorogo

Sugianti

Prodi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Ponorogo

PKH is one of the programs for providing conditional social assistance to Poor Families which are designated as recipient families, PKH benefits with the aim of relieving the economy of those who are less fortunate. However, in its implementation, especially in the community, this program has not run well and optimally due to the unevenness of the PKH program. This is of course very influential on the welfare of the community where the programs that should be received by the poor have not been well realized. The concept of data mining can be used to provide solutions to problems at hand, the Naïve Bayes Method is a method which can predict an opportunity from previous experience. The variables used include; Disability, children under five, school children and the elderly, where this variable is a requirement of a family as PKH recipients with additional criteria classified as a poor family with variables of occupation, income and residence status. Then the variable data is processed in the R language to determine the prediction results on the eligibility of PKH recipient families. The results of this study can be used as benchmarks and help in the optimization of the PKH program, from the experiments conducted, the results obtained were two households with id Test24 and Test44 which had the status of a recipient that was predicted to be non-recipient, with an Accuracy of 96%, 100% Sensitivity, and Specivicity 92%.

Pendahuluan

Kesejahteraan masyarakat merupakan salah satu cita-cita pemerintahan dalam menciptakan masyarakat yang makmur, adil dan beradab berdasarkan cita-cita yang tertulis di UUD 1945. Akan tetapi jumlah penduduk masyarakat Indonesia yang banyak menimbulkan beberapa permasalahan dalam pemerataan kehidupan yang layak, pemerintah terus melakukan pemerataan terhadap bantuan yang diberikan dari perkotaan maupun pedesaan¹.

PKH merupakan salah satu program bantuan pemerintah yang bertujuan untuk mengurangi angka kemiskinan masyarakat Indonesia yang ditujukan untuk Keluarga Penerima Manfaat (KPM). Bantuan program Keluarga Miskin (PKH) merupakan salah satu program bantuan kepada masyarakat kurang mampu adalah bantuan Program Keluarga Harapan yang merupakan salah satu bentuk bantuan yang ditujukan untuk memperkecil angka kemiskinan atau program bantuan bersyarat yang disalurkan untuk keluarga miskin yang diberlakukan sejak tahun 2007 oleh pemerintah Indonesia². Tujuan dari bantuan PKH untuk masyarakat kurang mampu dilihat berdasarkan data pada bulan Maret tahun 2016 angka kemiskinan Indonesia memiliki presentase 10,86% dari jumlah masyarakat Indonesia 28,01 juta jiwa. Pemerintah sendiri telah berupaya untuk menekan angka kemiskinan menurun 7-8% dari data yang diperoleh pada tahun 2019³.

Akan tetapi dalam prosesnya PKH memiliki beberapa kendala dalam pemerataan penyaluran bantuan tersebut, hal ini tentu saja sangat merugikan pemerintah dalam upaya penanggulangan

kemiskinan terutama bagi masyarakat kurang mampu isu dibuktikan dengan berita yang dilansir pada situs Detik.com pada januari 2019 tentang masyarakat miskin yang belum masuk data penerima PKH.

Untuk menanggulangi permasalahan tersebut diperlukanya sebuah penelitian untuk mengkaji ulang penyaluran bantuan PKH tersebut, Data mining adalah sebuah ilmu komputer yang menggunakan pembelajaran komputer untuk melakukan analisa dan mendapatkan informasi secara otomatis^{4 5}, KDD atau biasa disebut dengan *Knowledge discovery in databases* adalah salah satu penerapan dalam bidang data mining. Data Mining merupakan sebuah proses pengumpulan data yang berukuran besar atau *Big Data* kemudian mengekstraksi data tersebut menjadi informasi - informasi yang nantinya dapat digunakan untuk sebuah keperluan dengan cara mencari ciri dari suatu data tersebut^{6 7}.

Metode Klasifikasi adalah sebuah metode yang memanfaatkan sekumpulan data untuk mencari sebuah model atau fungsi yang bertujuan untuk memprediksi sebuah kelas tertentu yang hasilnya belum diketahui dari sebuah objek⁸. *Naïve Bayes Classifier* dikenal sebagai pengklasifikasi *bayesian* sederhana dan telah menjadi model probabilistik yang penting dan telah berhasil dalam praktiknya, *Naïve Bayes* memberikan rekomendasi sebaik model faktorisasi matriks, dan prediksi yang dapat dijelaskan^{9 10}. Algoritma *Naïve Bayes* menggunakan dua bentuk data dalam memulai prediksinya, dimana terdapat sebuah dataset dan datatest. Dataset adalah dimana sekumpulan data sebagai data latih yang digunakan untuk menentukan peluang yang akan terjadi¹¹, sedangkan datatest adalah sekumpulan data yang akan di prediksi dengan data latih yang sebelumnya telah dibuat dengan perhitungan probabilitas dan statistic.

Bahasa R merupakan sebuah bahasa pemrograman untuk analisis dan statistik yang bersifat *Open Source*. Bahasa R memiliki kelebihan antara lain memiliki kemampuan membuat grafis yang canggih, Sintaksnya mudah dipelajari dengan banyak fungsi-fungsi statistik yang terpasang¹². Bahasa R memiliki IDE yang sering dikenal sebagai R studio, dimana user lebih mudah dalam melakukan pengoperasian yang dilengkapi dengan *packages - packages* pendukung dalam proses menyelesaikan permasalahan ¹³, ¹⁴.

Dari beberapa permasalahan diatas penulis membuat sebuah penelitian tentang prediksi penerimaan bantuan PKH dengan menggunakan algoritma *Naïve Bayes* dengan menggunakan bahasa pemrograman R.

Metode Penelitian

Jenis dan Sumber data

Untuk memperoleh informasi dan data yang digunakan dalam penelitian ini adalah data primer dan data sekunder. Data primer yang digunakan dalam penelitian ini antara lain adalah wawancara dan observasi yang bertujuan untuk menentukan variable pendukung yang digunakan untuk proses prediksi penerima bantuan Program Keluarga Harapan (PKH) pada Desa Selur, Kecamatan Ngrayun, Kabupaten Ponorogo. Data sekunder yang digunakan adalah data penduduk desa berupa data pekerjaan penghasilan dan anggota keluarga dengan jumlah 200 kepala keluarga yang diambil dari data 5 tahun terakhir dan data penerima PKH gelombang 6 pada Desa Selur, Kecamatan Ngrayun, Kabupaten Ponorogo. Data tersebut dibagi menjadi dua jenis data yaitu data *Training* dan data *Test* kemudian diolah dan disimpan dalam file berformat .CSV.

Analisis Data

Dalam penelitian ini penulis menganalisis data yang bersangkutan dengan penerimaan bantuan Program Keluarga Harapan (PKH), Data yang didapatkan dari hasil penelitian dipergunakan untuk pembentukan data Training dan data Test, data Training digunakan untuk pemodelan dengan

algoritma *Naive Bayes*. Data yang digunakan untuk di proses pada tahap selanjutnya menggunakan atribut seperti pada [Table 1](#), [Table 2](#)

Table 1. *Tabel Atribut*

Table 2. *Predikat Penghasilan*

Predikat penghasilan diatas berdasarkan penggolongan pendapatan menurut Badan Pusat Statistik (BPS).

Hasil dan Pembahasan

Pengumpulan Data

Pada tahap pengumpulan data dilaksanakan pada minggu ke-2 bulan juni 2020, dengan didampingi sekertaris desa pengambilan data didapatkan dari data Desa Selur, Kecamatan Ngrayun, Kabupaten Ponorogo yang merupakan data real. Dengan data yang didapatkan berupa data

keluarga dan data penerima bantuan PKH.

Untuk langkah lebih lanjut penulis menggunakan aplikasi Microsoft Excel sebagai aplikasi pengolah data, dan menyeleksi data-data yang tidak diperlukan dalam penelitian. Hasil akhir yang didapatkan dari penelitian adalah algoritma Naïve Bayes memberikan akurasi sebanyak 96% dalam menyelesaikan permasalahan, kemudian memberikan prediksi terhadap data Test yang diuji.

Pembahasan

Data Selection

Data yang digunakan dalam kasus ini adalah data masyarakat di Desa Selur. Penyeleksian data dilakukan untuk mempermudah dalam proses data mining. Untuk atribut yang digunakan adalah Pekerjaan, Penghasilan, Status Tempat Tinggal, Disabilitas, Anak Balita, Anak Sekolah dan Lansia

Preprocessing

Pada tahapan preprocessing data dikelompokkan berdasarkan status dari data tersebut penerima atau bukan penerima. Dengan pengelompokan data ini dimaksudkan untuk mempermudah dalam pembuatan data training dan data test untuk perhitungan selanjutnya.

Transformasi

Dalam tahap transformation data dari tahapan preprocessing dapat digunakan untuk penyusunan sebuah data training. Data training merupakan data yang nantinya digunakan dalam pembuatan model perhitungan dalam prediksi menggunakan algoritma *naïve bayes*. Data training seperti pada [Table 3](#) didapatkan dari penggabungan data penerima bantuan dan bukan penerima bantuan, dimana dari data tersebut berisikan atribut-atribut yang relevan dan mendukung dalam menentukan hasil prediksi nantinya. Dalam kasus ini penulis memperoleh 200 jumlah data yang nantinya akan dibagi menjadi 150 data training dan 50 data test.

Table 3. *Data Training*

Implementasi

Implementasi antar muka aplikasi menggunakan packages Shiny, Shiny merupakan antar muka atau UI dari bahasa R, shiny merupakan antar muka berbasis web dimana aplikasi ini berjalan dengan pemrograman dan algoritma bahasa R. Untuk tutorial dan pengertian lebih lanjut mengenai packages Shiny bisa di kunjungi di situs resmi *Shiny.Rstudio.com*.

Packages pendukung untuk pembuatan antar muka dan packages yang digunakan untuk pemodelan naïve bayes adalah sebagai berikut:

1. Shiny
2. Shiny Dashboard
3. e1071
4. Caret

Use Case Diagram Aplikasi

Use Case merupakan suatu penggambaran dari fungsi yang berjalan dalam *interface* atau antar muka sistem. *Use Case* yang digunakan dapat dilihat pada [Figure 1](#)

Figure 1. *UseCase Diagram*

Berdasarkan gambar diatas pengguna dapat mengakses enam menu dalam sistem, akses menu home, panduan, input data *training* dan data *test* serta melakukan prediksi dan analisis. Pengguna dapat menghapus dan menginputkan data training dan data testing baru.

Tampilan Aplikasi

Pada halaman home seperti pada [Figure 2](#) menampilkan informasi tentang antar muka aplikasi, dan algoritma yang digunakan untuk proses analisis dan prediksi data test.

Figure 2. *HalamanHome*

Menu panduan seperti pada [Figure 3](#) merupakan salah satu langkah awal dalam penggunaan sistem, dimana pada menu ini terdapat beberapa panduan-panduan untuk penggunaan sistem. Panduan berisikan kebutuhan data, jenis data dan tahapan prediksi.

Figure 3. *Halaman Panduan*

Menu data training seperti pada [Figure 4](#) digunakan untuk memasukan atau input file data latih dalam suatu kasus yang akan di prediksi atau dianalisis file yang di input berjenis .CSV, data training yang digunakan pada kasus ini berjumlah 150 data. Dengan hasil sebagai berikut

Figure 4. Halaman Training

Menu data test seperti pada [Figure 5](#) digunakan untuk memasukan atau input file data test atau data uji dalam suatu kasus, file yang di input berjenis .CSV, data test yang digunakan pada kasus ini berjumlah 50 data. Dengan hasil sebagai berikut:

Figure 5. Halaman Training

Setelah memasukan data training dan data test, selanjutnya adalah proses data mining seperti terlihat pada [Figure 6](#). Dimana dalam tahapan ini peneliti menggunakan algoritma *Naïve Bayes* untuk melihat *output* dari dari status penerimaan bantuan PKH, selain itu menampilkan analisis dengan *Confusionmatrix*

Figure 6. Halaman Hitung

Untuk hasil yang di dapatkan dapat di prediksi dua kk dengan id Test24 dan Test44 yang mempunyai status Penerima yang di prediksi menjadi Bukan Penerima, dengan akurasi sebanyak 96%, Sensitivitas 100%, dan Spesifisitas 92%.

Pengujian Sistem

Tahap pengujian dalam pembentukan suatu sistem merupakan salah satu hal yang wajib untuk dilakukan. Pengujian ini bertujuan untuk memastikan sistem yang dibuat untuk memastikan sistem berjalan sesuai yang diinginkan, dan layak untuk digunakan. Dalam penelitian ini peneliti menggunakan sebuah metode pengujian sistem dengan alur kerja melihat masukan dan keluaran dari sebuah sistem yaitu metode *blackbox*. Dengan kata lain metode ini tidak melakukan pengujian dengan mempertimbangkan hasil input dan output pada sistem apakah output dari sistem tersebut memberikan hasil yang sesuai.

Dalam melakukan pengujian sistem memerlukan beberapa langkah pengujian, langkah tersebut dilakukan untuk melihat hasil pengujian. Langkah dan hasil pengujian dapat dilihat di [Table 4](#) .

Table 4. Langkah Dan Hasil Pengujian

Kesimpulan

Dari penelitian yang telah dilakukan, dengan menggunakan pemodelan prediktif dengan algoritma *naïve bayes* terhadap data penerima dan bukan penerima bantuan PKH yang diolah berdasarkan konsep *Data Mining* dan perhitungan *Naïve Bayes*, ditemukan dua data Test dengan kode KK Test 24 dan 44 Dengan Pekerjaan Wiraswasta, Penghasilan Cukup, Tempat Tinggal Hak Milik, Tidak Mempunyai Anak Balita, Tidak Memiliki Anggota Keluarga Disabilitas, Memiliki Anak Sekolah Dan Tidak Memiliki Keluarga Lansia, yang memiliki status Penerima mendapat rekomendasi prediksi sebagai Bukan penerima. Berdasarkan percobaan yang dilakukan akurasi yang didapatkan algoritma *Naïve Bayes* sebesar 96%, Sensitivitas sebesar 1,000 dan Spesifisitas sebesar 0,9259.

Dari kesimpulan diatas dapat perancangan sistem dapat dikembangkan dimana untuk kedepanya proses data mining dapat menggunakan lebih dari satu algoritma, dengan tujuan untuk memaksimalkan proses prediksi dan analisis dengan menggunakan algoritma yang memiliki akurasi tertinggi.

References

1. Bastian, A., & Bakhtiar, N. P. A. (2018, October). RANCANG BANGUN APLIKASI SISTEM PENDUKUNG KEPUTUSAN PENENTUAN MASYARAKAT MISKIN MENURUT KRITERIA KEMENSOS (Studi Kasus: Kelurahan Tonjong). In Prosiding Industrial Research Workshop and National Seminar (Vol. 9, pp. 280-287).
2. Bharata, H. K. and Sulistyowati, H. S. (2018). Analisa Sentimen Reuni 212 Menggunakan Bahasa R. Journal Informatics, Science & Technology 8.
3. Budi, S. (2007). Data Mining, Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Graha Ilmu (Yogyakarta).
4. Dr, Budiharto, S., Si, M., Kom, Rachmawati, S. R. N., Si, M., et al. (2013). Pengantar praktis

- pemrograman r untuk ilmu komputer (Halaman Moeka Publishing).
5. Efendi, H. (2015). Penerapan Arima Ensemble dalam Peramalan Kebutuhan Makanan Pokok di Kabupaten Jember (Doctoral dissertation. UNIVERSITAS MUHAMMADIYAH JEMBER).
 6. Fadlan, C., Ningsih, S., and Windarto, A. P. (2018). PENERAPAN METODE NAÏVE BAYES DALAM KLASIFIKASI KELAYAKAN KELUARGA PENERIMA BERAS RASTRA. doi: 10.32767/jutim.v3i1.286. <https://dx.doi.org/10.32767/jutim.v3i1.286>.
 7. Fajar, A. H. (2013) (Yogyakarta.: Andi Offset).
 8. Jumeilah, F. S. (2018). Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE dengan Naïve Bayes. doi: 10.21456/vol8iss1pp92-98. <https://dx.doi.org/10.21456/vol8iss1pp92-98>.
 9. Nusa, D. C. P. B. S. (2016).
 10. Purwanto, S. A., Sumartono, S., and Makmur, M. (2013). Implementasi Kebijakan Program Keluarga Harapan (PKH) Dalam Memutus Rantai Kemiskinan (Kajian di Kecamatan Mojosari Kabupaten Mojokerto). WACANA. Jurnal Sosial dan Humaniora 16, 79-96.
 11. Rahman, A. A. and Kurniawan, Y. I. (2018). Aplikasi Klasifikasi Penerima Kartu Indonesia Sehat Menggunakan Algoritma Naïve Bayes Classifier. Jurnal Teknologi dan Manajemen Informatika 4.
 12. Ridwan, M., Suyono, H., and Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. jurnal EECCIS 7, 59-64.
 13. Sari, I. M., Hasibuan, N. A., Silalahi, N. (2017). Implementasi, Mining, Underwear. Informasi dan Teknologi Ilmiah (INTI) 5.