

Pendekatan Machine Learning dalam Memprediksi Keluarga Penerima Program PKH

Irwan Agus Sobari¹, Robi Aziz Zuama²

¹Universitas Nusa Mandiri

²Universitas Bina Sarana Informatika

irwan.igb@nusamandiri.ac.id

robi.rbz@bsi.ac.id

Diterima	Direvisi	Disetujui
17-10-2022	18-11-2022	18-01-2023

Abstrak - Masalah kemiskinan di Indonesia masih menjadi fokus utama pemerintah dalam menetaskannya, program keluarga harapan (PKH) menjadi program prioritas pemerintah dalam upaya memberantas kemiskinan di Indonesia, fokus utama PKH adalah memberikan bantuan kepada Rumah Tangga Sangat Miskin (RTSM) untuk bisa mengakses pendidikan, kesehatan dan kesejahteraan sosial. Dalam menentukan keluarga yang berhak menerima bantuan PKH sering mengalami masalah, seperti kurang tepat sasaran dalam menentukan RTSM, ini di dasarkan kepada kelalaian petugas sehingga kurang akurat dalam validasi data yang banyak. Sistem otomatis yang dapat memprediksi RTSM dapat menjadi solusi atas permasalahan ini, sistem yang didasarkan pada model *machine learning*. Penelitian ini bertujuan untuk menganalisis model *machine learning Decision Tree (DT)*, *Support Vector Machine (SVM)*, *Naive Bayes (NB)* dan *Logistic Regression (LR)* dalam memprediksi RTSM yang akurat. Hasil menunjukkan bahwa *Logistic Regression* menjadi model yang optimal untuk di implementasikan dengan nilai AUC sebesar 0,999.

Kata Kunci: PKH, *Machine Learning*, *Naive Bayes*

Abstract - The problem of poverty in Indonesia is still the main focus of the government in incubating it, the Family Hope Program (PKH) is the government's priority program in efforts to eradicate poverty in Indonesia, the main focus of PKH is to assist Very Poor Households (RTSM) to be able to access education, health and education. social welfare. In determining the families who are entitled to receive PKH assistance, they often experience problems, such as not being right on target in determining the RTSM, this is based on the negligence of the officer so that it is less accurate in validating a lot of data. An automated system that can predict RTSM can be a solution to this problem, a system based on machine learning models. This study aims to analyze the machine learning model *Decision Tree (DT)*, *Support Vector Machine (SVM)*, *Naive Bayes (NB)* and *Logistic Regression (LR)* in predicting accurate RTSM. The results show that *Logistic Regression* is the optimal model to be implemented with an AUC value of 0.999.

Keywords: PKH, *Machine Learning*, *Naive Bayes*

PENDAHULUAN

Masalah utama yang masih dihadapi Indonesia adalah kemiskinan, ini menjadi perhatian penting bagi pemerintah Indonesia untuk segera menetaskannya dengan cepat dan tepat (Wintana et al., 2019). Beberapa upaya dilakukan untuk menanggulangi masalah kemiskinan ini, seperti pada

tahun 2007 pemerintah Indonesia mencanangkan program yang diharapkan memutus mata rantai kemiskinan yaitu Program Keluarga Harapan (PKH) yang bertujuan selain menetaskan kemiskinan, juga sebagai program perlindungan sosial dengan memberikan bantuan kepada Rumah Tangga Sangat Miskin (RTSM) di bidang pendidikan, kesehatan dan kesejahteraan sosial dalam meningkatkan



kesejahteraan dan kualitas sumber daya manusianya (Suleman & Resnawaty, 2017).

Faktanya di lapangan tidak sejalan dengan apa yang dicanangkan, bantuan PKH ini masih kurang tepat sasaran (Sofianto, 2020), ini disebabkan oleh kurang akuratnya validasi data, data yang sangat banyak menyulitkan petugas dalam memverifikasi data yang direkomendasikan untuk mendapatkan bantuan, masalah tersebut berdampak pada penerima bantuan PKH, keluarga yang harusnya berhak mendapat bantuan, justru tidak mendapatkannya (Pertiwi et al., 2019).

Beberapa penelitian sebelumnya memberikan solusi atas permasalahan ini, seperti membuat sistem pendukung keputusan Menggunakan Metode *Simple Additive Weighting* (Pertiwi et al., 2019), Logika Fuzzy Mamdani (Kurniadi et al., 2022), *fuzzy saw* (Rosmania & Sutikno, 2017) dan *analytic network process* (Izzah, 2019). Dan mengimplementasikan sistem berbasis algoritma *machine learning* untuk memprediksi penerima bantuan PKH *Decission Tree* (Bahtiar & Silitonga, 2020), Algoritma C.4.5 dan *naive bayes* (Fitriani, 2020) dan KNN (Fajriana, 2022). *Machine learning* di implementasikan yang memungkinkan komputer belajar dan mendapatkan pengetahuan berupa pola dari data-data masa lalu, dengan ini komputer dapat mengidentifikasi dan memahami data masukan, sehingga dapat membuat keputusan sendiri dengan berdasarkan data masukan (Zhou, 2021), (Hart et al., 2021).

Pada penelitian ini, kami mengimplementasikan *machine learning* dalam memprediksi penerima bantuan PKH dari kumpulan data penerima bantuan sosial PKH (SA'ADAWIYAH, 2020), untuk memahami pola proses penemuan pengetahuan pada keluarga yang berhak menerima bantuan. Dalam penelitian ini, empat model seperti *Naive Bayes*, *Logistic Regression*, *Support Vector Machine*, dan *Decission Tree* di evaluasi menggunakan software orange data mining dan bandingkan hasil kinerjanya, dalam memprediksi keluarga yang berhak menerima bantuan PKH.

METODOLOGI PENELITIAN

Dataset diambil dari penelitian sebelumnya yang dilakukan oleh (SA'ADAWIYAH, 2020). Dalam dataset ini terdapat 95 kasus yang diklasifikasikan menjadi dua kategori, yaitu miskin untuk mereka yang berhak menerima PKH dan tidak miskin bagi mereka yang tidak berhak menerima PKH. Data berisi 5 atribut seperti aset, bantuan lain, pekerjaan, penghasilan dan tabungan yang berisi informasi tentang kriteria penilaian. Tabel 1 menjelaskan deskripsi lengkap dari dataset.

Dataset dianalisis menggunakan perangkat lunak orange data mining (tersedia di <https://orangedatamining.com/>). Algoritma *machine learning* di implementasikan untuk melakukan tugas klasifikasi data biner pada dataset, untuk memprediksi apakah termasuk kategori miskin sebagai penerima PKH atau tidak miskin, penelitian ini telah mengusulkan empat algoritma *machine learning* yang berbeda seperti *Naive bayes*, *Logistic Regression SVM*, dan *Decission Tree*.

Tabel 1. Dataset description and characteristics

Attribute Name	Attribute Description	Range
Aset	Apakah mempunyai aset sendiri	Ya, Tidak
Bantuan Lain	Penerima Bantuan Lain Selain PKH	Ya, Tidak
Pekerjaan	Jenis Pekerjaan	Buruh Lepas, Karyawan, PNS/TNI/POLRI, Usaha Sendiri
Penghasilan	Besarnya Penghasilan bulanan	< 1.000.000 1.000.000 – 2.000.000 > 2.000.000
Tanggung	Jumlah Tanggungan dari penghasilan	> 2 <= 2
Hasil	Variabel kelas	Terima Tidak

HASIL DAN PEMBAHASAN

Implementasi model *machine learning* dapat membantu dinas sosial dalam menentukan keluarga yang berhak menerima bantuan dari program keluarga harapan (PKH) secara tepat dan akurat, sehingga dapat meningkatkan harapan hidup mereka. Algoritma *machine learning* telah di implementasikan dalam mengembangkan model prediksi penerima program PKH. Semua teknik klasifikasi diuji dan di evaluasi menggunakan aplikasi orange data mining. Data dibagi menjadi dua bagian, yaitu data train dan data test dengan masing-masing bagian data train 80% dan data tese sebesar 20%. Model *machine learning* seperti *Naive bayes*, *Logistic Regression SVM*, *Decission Tree* dan *Random Forest* dengan parameter yang telah ditentukan telah di diuji dalam memprediksi apakah termasuk keluarga penerima PKH atau tidak. Untuk mengukur hasil kinerja setiap model prediksi, maka diukur dengan berdasarkan parameter *area under curve* (AUC), *accuracy*, *F1 Score*, *precision*, dan

recall dan *confusion matrix*. 10-K *Fold Validation* di gunakan untuk menghindari masalah *under fitting* dan *over fitting*,

Dari tabel menunjukkan hasil kinerja dari semua model dengan menggunakan parameter, dari tabel tersebut di jelaskan AUC pada model *decission tree* sebesar 0,987, SVM 0,990, *naive bayes* 0,992 dan *logistic regression* sebesar 0,999 dan akurasi dari model yang di usulkan seperti *decission tree*, SVM, dan *Logistic regression* masing-masing adalah 0,989 dan untuk *naive bayes* akurasinya lebih kecil dibandingkan dengan model lain yaitu 0,968.

Tabel 2. *Performance evaluation*

Model	AUC	Accu racy	F1	Precisi on	Recall
Tree	0,987	0,989	0,989	0,990	0,989
SVM	0,990	0,989	0,989	0,990	0,989
Naive Bayes	0,992	0,968	0,968	0,970	0,968
Logistic Regres sion	0,999	0,989	0,989	0,990	0,989

Precision digunakan untuk mengukur kinerja model dalam memprediksi positif yang benar dari kumpulan data, berdasarkan tabel *precision* pada model *decision tree*, SVM, *Naive Bayes*, *Logistic Regression* masing-masing 0,990, 0,990, 0,970 dan 0,990.

Recall atau sensitivitas digunakan dalam mengukur kinerja model dalam proporsi kasus positif aktual penerima bantuan yang identifikasi benar oleh model. *Recall* berdasarkan tabel menunjukan pada masing-masing model *decissio tree* sebesar 0,989, SVM sebesar 0,989, *naive bayes* sebesar 0,968, dan LR sebesar 0,989.

Secara singkat, F1 score adalah analisis penggabungan kinerja precision dan recall menjadi satu metrik dengan menghitung rata-rata antara keduanya, semakin mendekati 1, maka nilai F1 score disebut kinerja model terbaik. Berdasarkan tabel kinerja model DT, SVM, NB dan LR masing-masing 0,989, 0,989, 0,968 dan 0,989.

Tabel 3 dan tabel 4 menunjukan hasil *confution matrix* dari hasil kinerja model yang diusulkan.

Tabel 3. *Confution Matrix DT, SVM & LR*

		Predicted		
		Terima	Tidak	Σ
Actual	Terima	39	1	40
	Tidak	0	55	55
	Σ	39	56	95

Tabel 4. *Confution Matrix NB*

		Predicted		
		Terima	Tidak	Σ
Actual	Terima	37	3	40
	Tidak	0	55	55
	Σ	37	58	95

Jadi, dari hasil penelitian, dapat dikatakan bahwa berdasarkan semua parameter yang digunakan, semua model berkinerja baik dalam memprediksi keluarga yang berhak menerima bantuan PKH, akan tetapi, dari semua model yang diusulkan hanya model *naive bayes* yang kerjanya lebih rendah dibandingkan semua model. Jika hanya memperhitungkan akurasi kinerja model saja, maka kinerja akurasi tidak memberikan banyak wawasan yang lebih baik, F1 Score memberikan wawasan yang baik, karena mampu menjelaskan antara *precision* dan *recall*. Nilai AUC *logistic regression* menunjukkan bahwa model berkinerja terbaik sebesar 0,999 hampir mendekati 1 dapat dinyatakan bahwa LR adalah model pengklasifikasi yang optimal untuk memprediksi.

KESIMPULAN

Penelitian ini telah mengembangkan empat model pembelajaran machine learning seperti *Decision Tree* (DT), *Support Vector Machine* (SVM), *Naive Bayes* (NB) dan *Logistic Regression* (LR) dalam memprediksi keluarga penerima bantuan PKH.

Semua model di analisis menggunakan software orange data mining, dan di evaluasi menggunakan parameter AUC, *Accuracy*, F1 Score, *Precision* dan *Recall*. Hasil experiment menunjukkan kinerja model yang diusulkan sudah sangat baik. Sperti DT dengan hasil akurasi 0,989, SVM dengan hasil akurasi 0,989, NB dengan hasil akurasi 0,968 dan LR dengan hasil akurasi 0,989, namun dari hasil analisis NB menjadi model dengan kinerja terendah dibandingkan dengan hasil model yang lain. Nilai F1 Score memberikan wawasan lebih baik, semua model yang diusulkan mendapatkan nilai F1 Score yang baik, tapi untuk model *Naive Bayes* kinerja F1

score lebih rendah dibanding dengan kinerja model lain. Nilai AUC yang tinggi, menunjukkan kinerja optimal dari model, *logistic regression* menjadi model dengan nilai AUC tertinggi yaitu 0,999 dan menjadi model yang optimal dalam kinerja klasifikasi keluarga penerima PKH.

REFERENSI

- Bahtiar, A., & Silitonga, P. D. P. (2020). Penerapan Algoritma Decision Tree Untuk Memprediksi Penerima Bantuan Keluarga Harapan. *Jurnal ICT: Information Communication & Technology*, 19(1), 70–76.
- Fajriana, F. (2022). Classification of Determination the Recipients of the Program Keluarga Harapan (PKH) Using K-Nearest Neighbor Algorithm. *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, 6(1), 298–308.
- Fitriani, E. (2020). Perbandingan Algoritma C4. 5 Dan Naïve Bayes Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan. *SISTEMASI: Jurnal Sistem Informasi*, 9(1), 103–115.
- Hart, G. L. W., Mueller, T., Toher, C., & Curtarolo, S. (2021). Machine learning for alloys. *Nature Reviews Materials*, 6(8), 730–755.
- Izzah, K. (2019). *Sistem pendukung keputusan kelayakan penerima program keluarga harapan (pkh) menggunakan algoritma analytic network process*. UIN Sunan Ampel Surabaya.
- Kurniadi, D., Nuraeni, F., & Jaelani, D. (2022). Implementasi Logika Fuzzy Mamdani Pada Sistem Prediksi Calon Penerima Program Keluarga Harapan. *Jurnal Algoritma*, 19(1), 151–162.
- Pertiwi, I. P., Fedinandus, F. X., & Limantara, A. D. (2019). Sistem Pendukung Keputusan Penerima Program Keluarga Harapan (PKH) Menggunakan Metode Simple Additive Weighting. *CAHAYATECH*, 8(2), 182–195.
- Rosmania, F., & Sutikno, S. (2017). *SISTEM PENDUKUNG KEPUTUSAN SELEKSI PENERIMA PROGRAM KELUARGA HARAPAN (PKH) MENGGUNAKAN FUZZY SIMPLE ADDITIVE WEIGHTING (FUZZY SAW)*(Studi Kasus: Unit Pelaksana Program Keluarga Harapan Kecamatan Tembalang Kota Semarang). Universitas Diponegoro.
- SA'ADAWIYAH, S. (2020). *PENERAPAN ALGORITMA NAIVE BAYES DALAM PENENTUAN PENERIMA BANTUAN SOSIAL PROGRAM KELUARGA HARAPAN (PKH)*.
- Sofianto, A. (2020). Implementasi Program Keluarga Harapan (PKH) di Provinsi Jawa Tengah. *Sosio Konsepsia*, 10(1), 14–31.
- Suleman, S. A., & Resnawaty, R. (2017). Program Keluarga Harapan (PKH): Antara perlindungan sosial dan pengentasan kemiskinan. *Prosiding Penelitian Dan Pengabdian Kepada Masyarakat*, 4(1), 88–92.
- Wintana, D., Hikmatulloh, H., Ichsan, N., Purnama, J. J., & Rahmawati, A. (2019). Klasifikasi Penentuan Penerima Manfaat Program Keluarga Harapan (Pkh) Menggunakan Algoritma C5. 0 (Studi Kasus: Desa Sukamaju, Kec. Kadudampit). *KLIK-KUMPULAN JURNAL ILMU KOMPUTER*, 6(3), 254–263.
- Zhou, Z.-H. (2021). *Machine learning*. Springer Nature.