KLASIFIKASI RUMAH TIDAK LAYAK HUNI STUDI PERBANDINGAN 9 MODEL *MACHINE LEARNING*

Deriska Fadilla Musdalifa, Roni Andarsyah, Roni Habibi
Teknik Informatika, Vokasi, Universitas Logistik dan Bisnis Internasional
deriskafadillam18@gmail.com

ABSTRAK

Rumah tidak layak huni (RTLH) merupakan masalah serius yang dihadapi oleh banyak masyarakat di berbagai negara. Identifikasi rumah tidak layak huni secara efektif dapat membantu dalam upaya perbaikan dan perencanaan perumahan yang lebih baik. Dalam menentukan keluarga yang berhak menerima bantuan RTLH sering mengalami masalah, seperti kurang tepat sasaran dalam penentuan RTLH. Dalam era kemajuan teknologi, model machine learning telah muncul sebagai alat yang kuat untuk mengatasi berbagai masalah kompleks, termasuk identifikasi rumah tidak layak huni. Tujuan dari penelitian ini adalah untuk melakukan komparasi antara beberapa model *machine learning* yang paling umum digunakan dalam identifikasi rumah tidak layak huni. Model-model tersebut meliputi *K-Nearest Neighbors, Naïve Bayes, Support Vector Machine, pohon keputusan (Decision Tree), Logistic Regression, Random Forest, Neural Network, Gradient Boost, dan XGBoost*. Data yang digunakan dalam penelitian ini berisi atribut-atribut rumah yang relevan seperti penghasilan, jumlah penghuni rumah, luas rumah, kondisi struktur bangunan, sanitasi, serta listrik. Evaluasi dilakukan dengan membandingkan performa model-model tersebut berdasarkan metrik evaluasi yang relevan, seperti akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa model *Neural Network* menjadi model yang optimal untuk di implementasikan dengan nilai akurasi 99,24%, presisi 98,66%, recall 100% serta F1-score 99,32%.

***Kata kunci**: RTLH, Machine Learning, Neural Network.*

1. PENDAHULUAN

Rumah tidak layak huni merupakan masalah yang meresahkan banyak masyarakat di seluruh dunia [1]. Dalam konteks ini, rumah tidak layak huni dapat diartikan sebagai kondisi hunian yang tidak memenuhi standar kesehatan, keamanan, dan kelayakan minimum yang diperlukan untuk kehidupan manusia yang layak [2]. Fenomena ini memiliki dampak negatif yang signifikan terhadap kesejahteraan dan kualitas hidup penduduk di berbagai negara [3]

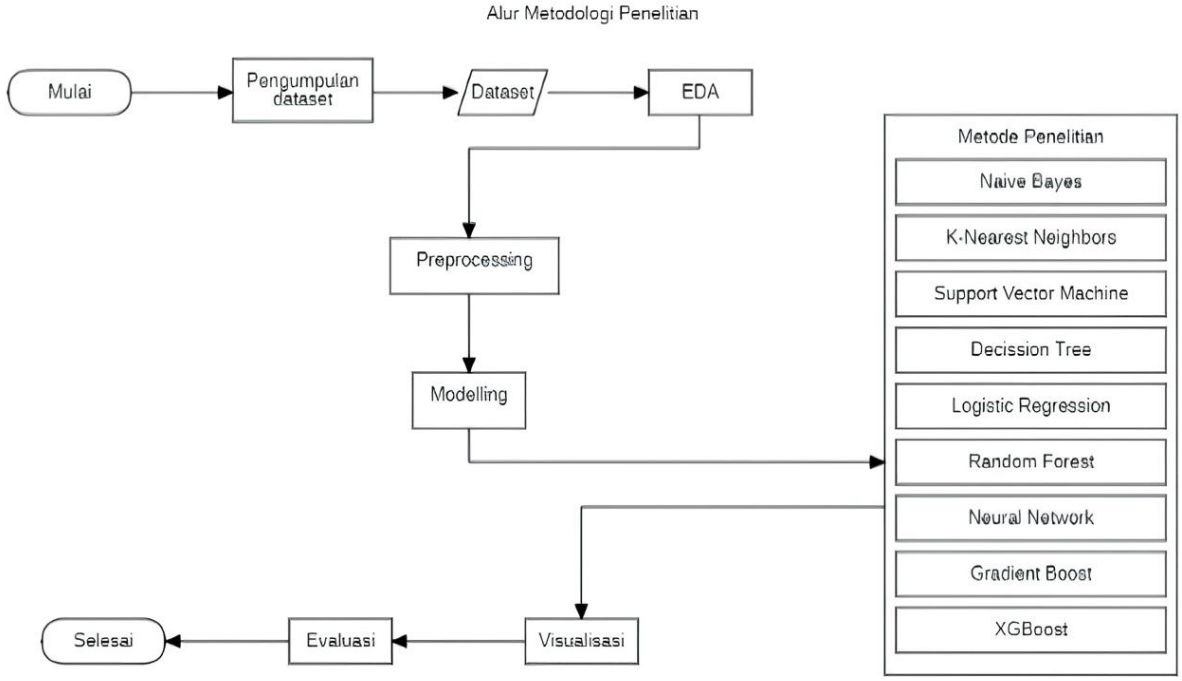
Dalam beberapa tahun terakhir, perkembangan teknologi dan penggunaan machine learning telah memberikan potensi baru dalam upaya mengatasi rumah tidak layak huni. *Machine learning* merupakan pendekatan yang memungkinkan analisis data secara otomatis dan pengambilan keputusan berdasarkan pola dan informasi yang terdapat dalam data [4].

2. TINJAUAN PUSTAKA

Beberapa penelitian terdahulu telah melakukan penelitian tentang klasifikasi bantuan sosial menggunakan *machine learning* dengan berbagai metode [5]. Penelitian [6] menggunakan metode data mining C4.5 dan software *RapidMiner* untuk melakukan klasifikasi bantuan sosial. Mereka mencapai akurasi sebesar 83,33% menggunakan metode *validasi split*. Penelitian yang dilakukan oleh [7] menggunakan metode KNN mendapatkan hasil akurasi sebesar 96,25%. Penelitian lain oleh [8] penggunaan metode klasifikasi *Naïve Bayes* dengan 50 data latih dan 10 data uji, penelitian ini mencapai akurasi sebesar 80%. Pada penelitian [9] menggunakan 160 data yang terbagi menjadi dua kategori yaitu layak dan tidak layak menggunakan metode SVM. Hasil pengujian menunjukkan rata-rata akurasi sebesar 98,75% dengan menggunakan metode *K-fold Cross Validation* dengan nilai k = 10. Komparasi model SVM (*Support Vector Machine*), NB (*Naïve Bayes*), dan RF (*Random Forest*) oleh [10] yang dievaluasi menggunakan metrik akurasi, presisi, *recall* dan *f-measure* mendapatkan hasil 97,26% pada RF dengan perbedaan sebesar 5,11% dengan algoritma SVM dan 8,87% pada klasifikasi NB. Perbandingan algoritma C4.5 dan NB pada penelitian [11] pada program penerima keluarga harapan C4.5 menghasilkan tingkat akurasi sebesar 91,25% dan NB menghasilkan 87,11%.

Kontribusi dari penelitian-penelitian tersebut adalah menyediakan metode-metode yang berbeda untuk melakukan klasifikasi bantuan sosial menggunakan *machine learning*. Dengan mencapai tingkat akurasi yang signifikan, penelitian-penelitian ini dapat menjadi landasan untuk pengembangan dan implementasi sistem klasifikasi yang lebih efektif dalam konteks bantuan sosial. Berdasarkan beberapa penelitian diatas, maka penulis membuat komparasi model menggunakan *machine learning* untuk memprediksi keluarga yang berhak menerima bantuan Rumah Tidak Layak Huni (RTLH). Penulis mengevaluasi sembilan model, termasuk *KNN, Naive Bayes, SVM, Decision Tree, Linear Regression, Random Forest, Neural Network, Gradient Booster, dan XGBoost,* menggunakan metode data mining untuk membandingkan kinerja terbaik dalam memprediksi keluarga yang berhak menerima bantuan RTLH pada data yang telah diperoleh.

3. METODE PENELITIAN



Gambar 1. Metode Penelitian

Melalui alur metode penelitian ini, dapat dikembangkan model *machine learning* yang efektif dalam mengidentifikasi rumah tidak layak huni berdasarkan atribut-atribut yang relevan. Pendekatan ini memungkinkan pemanfaatan berbagai algoritma dan teknik yang sesuai untuk mencapai tujuan identifikasi yang akurat dan efisien.

3.1. Pengumpulan Data

Pengumpulan dataset diambil dari hasil survey di dinas PUPR kota Blitar. Dalam dataset ini terdapat 1703 data yang terdapat 25 kolom yaitu, No BNBA, Pekerjaan, Status Kepemilikan Tanah, Status Kepemilikan Rumah, Pernah Menerima Bantuan, Penghasilan, Jumlah KK dalam 1 Rumah, Jumlah Penghuni Rumah, Luas Rumah, Pondasi Rumah, Kondisi Sloof, Kondisi Kolom, Kondisi Balok, Kondisi Struktur Atap, Kondisi Atap, Kondisi Dinding, Kondisi Lantai, Material Lantai, Material Dinding, Material Atap, Kondisi Jendela /Pencahayaan, Kondisi Fentilasi, Kondisi Sumber Air Minum, Kondisi Jamban, serta Kepemilikan Listrik. Parameter yang dipakai hanya 20 kolom dan setelah itu diklasifikasikan menjadi dua kategori, yaitu layak huni dan tidak layak huni. Dapat dilihat pada tabel 1 dibawah untuk beberapa data yang diperoleh

Tabel 1. Dataset

No. BNBA	11006001	11006002	11006003	11006004	11006005
Pekerjaan	Buruh Harian/Tukang	WIRAUSAHA	TIDAK BEKERJA	Ojek/Supir	WIRAUSAHA
Status Kepemilikan Tanah	WARISAN	WARISAN	Milik Sendiri	Milik Sendiri	WARISAN
Status Kepemilikan Rumah	Milik Sendiri	Milik Sendiri	Milik Sendiri	Milik Sendiri	Milik Sendiri
Pernah Menerima Bantuan?	Belum Pernah	Belum Pernah	Belum Pernah	Pernah	Belum Pernah
Penghasilan	< 1,2 Juta	1,2 - 1,8 Juta	TIDAK BEKERJA	< 1,2 Juta	< 1,2 Juta
Jumlah KK dalam 1 Rumah	1	2	2	1	1
Jumlah Penghuni Rumah	2	2	3	5	2
Luas Rumah	24	48	48	32	32
Pondasi Rumah	1	1	4	1	4
Kondisi Sloof	1	3	5	2	6
Kondisi Kolom	2	3	5	2	6

No. BNBA	11006001	11006002	11006003	11006004	11006005
Kondisi Balok	2	3	5	2	6
Kondisi Struktur Atap	1	3	5	2	6
Kondisi Atap	1	3	5	1	4
Kondisi Dinding	1	1	4	2	4
Kondisi Lantai	3	1	4	2	4
Material Lantai	Plesteran	Plesteran	Plesteran	Plesteran	Plesteran
Material Dinding	Plesteran	Plesteran	Plesteran	Plesteran	Plesteran
Material Atap	Seng	Genteng	Genteng	Genteng	Genteng
Kondisi Jendela / Pencahayaan	Mencukupi	Mencukupi	Mencukupi	Mencukupi	Mencukupi
Kondisi Ventilasi	Mencukupi	Mencukupi	Mencukupi	Mencukupi	Mencukupi
Kondisi Sumber Air Minum	Sumur	Sumur	Sumur	Sumur	Sumur
Kondisi Jamban	BELUM Sama Sekali / MENUMPANG	Ada, Layak	Ada, Tidak Layak	Ada, Layak	Ada, Layak
Kepemilikan Listrik	Milik Sendiri	Milik Sendiri	Milik Sendiri	Milik Sendiri	MENUMPANG

3.2. Pengolahan Data

Preprocessing atau pengolahan data dalam metodologi penelitian sangat penting untuk mempersiapkan data sebelum analisis lebih lanjut. Salah satu aspek penting dalam tahap ini adalah *cleaning data*, yang bertujuan untuk memastikan keakuratan, konsistensi, dan kegunaan data dalam dataset [12]. Pada tahap c*leaning data*, dilakukan langkah-langkah untuk mengidentifikasi dan memperbaiki masalah dalam data, seperti *missing values*, *outliers*, atau data yang tidak sesuai format. Tujuan utamanya adalah memastikan data yang digunakan dalam analisis berkualitas, terpercaya, dan relevan, sehingga mengurangi bias dan meningkatkan efisiensi serta akurasi proses analisis selanjutnya [13]. Pada bagian pembobotan, dilakukan penggantian nilai-nilai kategori dalam kolom-kolom tertentu dengan angka-angka yang merepresentasikan tingkat atau kualitas tertentu. Misalnya, pada kolom 'Material Lantai', nilai kategori seperti 'keramik', 'plesteran', 'ubin', 'rabat', dan 'tanah' digantikan dengan angka 4, 3, 2, 1, dan 0 secara berturut-turut. Hal yang sama dilakukan pada kolom-kolom lainnya seperti 'Material Dinding', 'Material Atap', 'Kondisi Jendela / Pencahayaan', 'Kondisi Ventilasi', 'Kondisi Sumber Air Minum', 'Kondisi Jamban', 'Kepemilikan Listrik', dan 'Penghasilan'.

3.3. Pembagian Data

Pembagian data adalah langkah dalam pemrosesan data yang melibatkan pemisahan dataset menjadi subset yang berbeda. Tujuan dari pembagian data adalah untuk memisahkan data menjadi bagian yang dapat digunakan untuk melatih model (data latih) dan bagian yang digunakan untuk menguji performa model (data uji) [14].

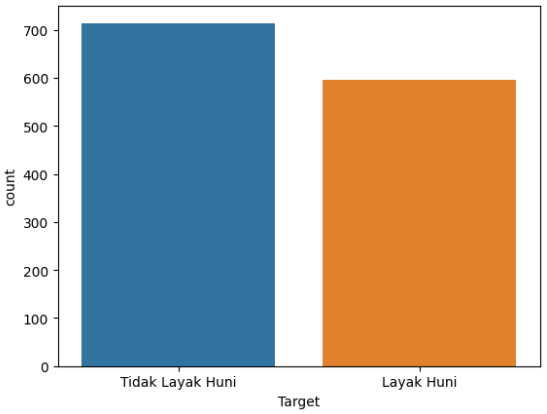
Parameter test_size=0.2 menentukan proporsi data yang akan dijadikan data uji, dalam hal ini 20% dari dataset akan digunakan sebagai data uji, sementara 80% sisanya akan digunakan sebagai data latih. Proporsi ini dapat diubah sesuai kebutuhan.

Parameter random_state=42 digunakan untuk mengatur seed atau bilangan acak yang digunakan dalam pembagian data. Dengan menggunakan nilai random_state yang sama, pembagian data akan konsisten setiap kali kode dieksekusi.

3.4. Pelabelan Data

Pelabelan data adalah proses memberikan label atau klasifikasi pada data berdasarkan suatu aturan atau kondisi tertentu. Tujuan dari pelabelan data adalah untuk mengidentifikasi dan membedakan kelompok atau kategori data yang berbeda [15].

Pelabelan data dilakukan pada kolom 'Target' berdasarkan nilai pada kolom 'Sum'. Nilai rata-rata ('avg_value') digunakan sebagai titik pemisah untuk memberikan label 'Layak Huni' atau 'Tidak Layak Huni'. Jika nilai pada kolom 'Sum' lebih besar atau sama dengan nilai rata-rata, maka data diberi label 'Layak Huni'. Jika nilai lebih kecil dari nilai rata-rata, data diberi label 'Tidak Layak Huni'.



Gambar 2. Hasil Pelabelan

Pelabelan data ini dapat membantu dalam analisis dan pemahaman lebih lanjut terhadap data. Dengan memberikan label pada data, kita dapat mengidentifikasi kelompok atau kategori tertentu, dan

melakukan pemodelan atau analisis statistik yang lebih spesifik terhadap masing-masing kelompok. Selain itu, pelabelan data juga mempermudah interpretasi dan visualisasi data dalam bentuk yang lebih mudah dipahami [16].

3.5. Pemodelan Data

Pemodelan data adalah proses menganalisis data secara sistematis. Tujuannya adalah untuk mengidentifikasi pola, hubungan, dan tren yang terdapat dalam data, serta membuat prediksi atau mengambil keputusan berdasarkan pemahaman tersebut [17]. Dengan memanfaatkan berbagai teknik seperti *K-Nearest Neighbors (KNN), Naive Bayes, SVM (Support Vector Machine), Decision Tree, Linear Regression, Random Forest, Neural Network, Gradient Booster, dan XGBoost*, pemodelan data memungkinkan kita untuk menggabungkan pendekatan yang berbeda dan memperoleh pemahaman yang lebih lengkap tentang data yang dianalisis. Metode-metode ini memainkan peran penting dalam berbagai bidang, seperti ilmu sosial, bisnis, ilmu pengetahuan alam, kesehatan, dan keuangan, dan membantu kita membuat keputusan yang lebih informasional dan prediksi yang lebih akurat berdasarkan data yang ada.

3.6. Evaluasi

Evaluasi dilakukan dengan membandingkan hasil prediksi model dengan data pengujian yang sudah diketahui kebenarannya [18]. Metrik evaluasi yang umum digunakan meliputi akurasi, presisi, recall, dan f1-score. Dengan evaluasi ini, kita dapat mengevaluasi sejauh mana model dapat memprediksi dengan akurat, menemukan kasus positif, dan memberikan keseluruhan performa yang baik. Evaluasi membantu kita memahami sejauh mana model yang telah dibangun efektif dan dapat dipercaya dalam mengambil keputusan atau membuat prediksi berdasarkan data yang ada [19].

4. HASIL DAN PEMBAHASAN

Dalam dataset, terdapat beberapa kolom yang memiliki nilai-nilai kategorikal yang digantikan dengan nilai numerikal. Misalnya, dalam kolom 'Pernah Menerima Bantuan?', nilai 'belum pernah' digantikan dengan nilai 1 dan nilai 'pernah' digantikan dengan nilai 0. Tujuannya adalah untuk mengubah representasi nilai agar lebih sesuai untuk analisis selanjutnya.

Hal yang sama juga dilakukan pada kolom 'Material Lantai', 'Material Dinding', dan 'Material Atap', di mana nilai kategorikal digantikan dengan nilai numerikal yang menggambarkan jenis material. Nilai numerikal yang lebih tinggi menunjukkan material yang lebih baik atau lebih mahal.

Setelah penggantian nilai selesai, dapat dilakukan perhitungan akurasi, presisi, recall, dan f1-score menggunakan model yang relevan dengan dataset tersebut. Hasil metrik-metrik tersebut dapat digunakan untuk membandingkan performa model-model tersebut. Untuk memvisualisasikan perbandingan model, dapat menggunakan *library visualisasi* data seperti *Matplotlib* atau Seaborn dengan menggunakan diagram batang. Dengan menggunakan visualisasi data, perbandingan antara akurasi, presisi, recall, dan f1-score dari model-model tersebut dapat dilihat dengan lebih jelas dan mudah dipahami.

4.1. Hasil Evaluasi

Tabel 2. Hasil Evaluasi

Model	Akurasi	Presisi	Recall	F1-Score
KNN	75.95%	78.00%	79.59%	78.79%
Naïve Bayes	74.81%	73.14%	87.07%	79.50%
SVM	95.42%	95.92%	95.92%	95.92%
Decission Tree	91.60%	92.52%	92.52%	92.52%
Logistic Regression	82.06%	82.47%	86.39%	84.39%
Random Forest	93.51%	92.76%	95.92%	94.31%
Neural Network	99.24%	98.66%	100%	99.32%
Gradient Boost	94.27%	93.42%	96.60%	94.98%
XGBoost	94.27%	93.42%	96.60%	94.98%

Hasil evaluasi dari berbagai model machine learning untuk mengklasifikasikan rumah tidak layak huni (RTLH) yaitu KNN (*K-Nearest Neighbors*) memiliki akurasi 76%, dengan presisi dan recall yang seimbang untuk kedua kelas. *Naïve Bayes* memiliki akurasi 75%, dengan presisi yang baik untuk kedua kelas, tetapi recall untuk kelas 0 lebih rendah. SVM (*Support Vector Machine*) memiliki performa yang sangat baik dengan akurasi 95%, presisi, dan recall yang tinggi untuk kedua kelas. *Decision Tree* memiliki akurasi 92%, dengan tingkat presisi dan recall yang tinggi untuk kedua kelas. *Logistic Regression* memiliki akurasi 82% dengan presisi dan recall yang baik untuk kedua kelas. *Random Forest* memiliki performa yang sangat baik dengan akurasi 94%, dan presisi serta recall yang tinggi untuk kedua kelas. *Neural Network* menunjukkan performa yang sangat baik dengan akurasi 99%, dan presisi serta recall yang tinggi untuk kedua kelas. *Gradient Boost* memiliki performa yang baik dengan akurasi 94%, dan presisi serta recall yang baik untuk kedua kelas. *XGBoost* juga menunjukkan performa yang baik dengan akurasi 94%, dengan presisi dan recall yang tinggi untuk kedua kelas.

Secara keseluruhan, hampir semua model machine learning menunjukkan hasil yang baik dalam mengklasifikasikan rumah tidak layak huni. Namun, model *Neural Network* mencapai akurasi tertinggi dengan nilai 99% dan presisi serta recall yang tinggi untuk kedua kelas. Model *SVM, Random Forest*, dan *XGBoost* juga memiliki performa yang sangat baik

dengan akurasi di atas 90%. Pemilihan model yang optimal dapat didasarkan pada preferensi dan kebutuhan penelitian serta perbandingan lebih lanjut terhadap metrik evaluasi yang relevan.

5. KESIMPULAN

Dalam penelitian ini, telah dibandingkan 9 model *machine learning*, yaitu *KNN, NaiveBaiyes, SVM, Decission Tree, Linear Regression, Random Forest, Neural Network, Gradient Booster, XGBoost*, untuk memprediksi keluarga penerima bantuan Program Rumah Tidak Layak Huni (RTLH). Semua model dianalisis menggunakan *Data Mining* pada *Google Collab* dan dievaluasi menggunakan parameter *Accuracy, Precision, Recall*, dan *F1 Score*. Jika diurutkan hasil menunjukkan KNN memiliki akurasi sebesar 75,95%, NB memiliki akurasi sebesar 74,81%, LR memiliki akurasi sebesar 82,06%, DT memiliki akurasi sebesar 91,60%, RF memiliki akurasi sebesar 93,51%, GB memiliki akurasi sebesar 94,27%, XGB memiliki akurasi sebesar 94,27%, SVM memiliki akurasi sebesar 95,42%, serta Neural Network memiliki akurasi tertinggi yaitu sebesar 99,24% yang menunjukkan kinerja optimal sehingga model *Neural Network* menjadi model optimal dalam klasifikasi Rumah Tidak Layak Huni.

DAFTAR PUSTAKA

[1] I. A. Sobari and R. A. Zuama, "This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License Jurnal Teknik Komputer AMIK BSI Pendekatan Machine Learning dalam Memprediksi Keluarga Penerima Program PKH", doi: 10.31294/jtk.v4i2.

[2] A. Naas, S. Na'iema, H. Mulyo, and A. Widiastuti, "Klasifikasi penerima bantuan program rehabilitasi rumah tidak layak huni menggunakan algoritme K-Nearest Neighbor Classification of beneficiaries for the rehabilitation of uninhabitable houses using the K-Nearest Neighbor algorithm," *Jurnal Teknologi dan Sistem Komputer*, vol. 10, no. 1, pp. 32–37, 2022, doi: 10.14710/jtsiskom.2022.14110.

[3] Irmayansyah and Aulia Arief Firdaus, "PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI PENENTUAN PENERIMAAN BANTUAN LANGSUNG DI DESA CIOMAS," *Jurnal Ilmiah Teknologi - Informasi dan Sains (TeknoIS)*, vol. 8, pp. 17–28, May 2018.

[4] Ilham Akhyar Firdaus, "Deteksi Infeksi Mycoplasma Pneumoniae Pneumonia Menggunakan Komparasi Algoritma Klasifikasi Machine Learning," 2018.

[5] F. Handayani *et al.*, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network dalam Prediksi Penyakit Jantung".

[6] A. Sugarda, A. Perdana Windarto, and W. Robiansyah, "Penerapan Metode Data Mining C4.5 dalam Penentuan Kelayakan Rehabilitas Rumah Warga," 2022.

[7] P. Algoritma K-Nearest Neighbor Untuk Klasifikasi Rumah Layak Atau Tidak Layak Huni et al., "PENERAPAN ALGORITMA K-NEAREST NEIGHBOR UNTUK KLASIFIKASI RUMAH LAYAK ATAU TIDAK LAYAK HUNI (STUDI KASUS: DESA BULU KECAMATAN KRAKSAAN KABUPATEN PROBOLINGGO)," Jurnal.ilmiah.informatika, vol. 7, no. 2, pp. 75–84, 2022, doi: 10.35316/jimi.v7i2.75-84.

[8] A. A. Saputro, "Sistem Pendukung Keputusan Penerimaan Bantuan Sosial Program Keluarga Harapan (PKH) dengan Menggunakan Metode Naïve Bayes Classifier (Studi Kasus di Balai Desa Bendungan Kraton Pasuruan)," Jurnal Ilmiah Edutic : Pendidikan dan Informatika, vol. 9, no. 1, pp. 40–48, Nov. 2022, doi: 10.21107/edutic.v9i1.12232.

[9] W. Agustina, M. T. Furqon, and B. Rahayudi, "Implementasi Metode Support Vector Machine (SVM) Untuk Klasifikasi Rumah Layak Huni (Studi Kasus: Desa Kidal Kecamatan Tumpang Kabupaten Malang)," 2018. [Online]. Available: http://j-ptiik.ub.ac.id

[10] I. Kurniawan, D. Cahya Putri Buani, W. Apriliah, R. Amegia Saputra, and P. Korespondensi, "IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK MENENTUKAN PENERIMA BANTUAN RASKIN," JTIIK, vol. 10, no. 2, pp. 421–428, 2023, doi: 10.25126/jtiik.202396225.

[11] Eka Fitriani, "PERBANDINGAN ALGORITMA C4.5DAN NAÏVE BAYES UNTUK MENENTUKAN KELAYAKAN PENERIMA BANTUAN PROGRAM KELUARGA HARAPAN," SISTEMASI, vol. 9, pp. 103–115, 2020.

[12] M. Ravly Andryan et al., "KOMPARASI KINERJA ALGORITMA XGBOOST DAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK DIAGNOSA PENYAKIT KANKER PAYUDARA," Jurnal Informatika dan Komputer), vol. 6, no. 1, pp. 1–5, 2022.

[13] L. Efrizoni, S. Defit, M. Tajuddin, and A. Anggrawan, "Komparasi Ekstraksi Fitur dalam Klasifikasi Teks Multilabel Menggunakan Algoritma Machine Learning," MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 21, no. 3, pp. 653–666, Jul. 2022, doi: 10.30812/matrik.v21i3.1851.

[14] Nuli Giarsyani, Ahmad Fathan Hidayatullah, and Ridho Rahmadi, "KOMPARASI ALGORITMA MACHINE LEARNING DAN DEEP LEARNING UNTUK NAMED ENTITY

RECOGNITION : STUDI KASUS DATA KEBENCANAAN," JIRE, vol. 3, 2020.

[15] C. Haryanto, N. Rahaningsih, and F. Muhammad Basysyar, "KOMPARASI ALGORITMA MACHINE LEARNING DALAM MEMPREDIKSI HARGA RUMAH," 2023.

[16] T. Herdiawan Apandi, C. Agus Sugianto, M. Informatika, and P. Negeri Subang, "ANALISIS KOMPARASI MACHINE LEARNING PADA DATA SPAM SMS," 2018.

[17] R. Nisa Sofia Amriza, D. Supriyadi, P. DI Jl Panjaitan No, K. Purwokerto Selatan, K. Banyumas, and J. Tengah, "Komparasi Metode Machine Learning dan Deep Learning untuk Deteksi Emosi pada Text di Sosial Media."

[18] Paul V.M., IndraGunawan, Bahrudi Efendi Damanik, Iin Parlina, and Widodo Saputra, "PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA C4.5 DALAM MENENTUKAN KELAYAKAN PENERIMAAN BANTUAN BEDAH RUMAH PADA DESA TIGA DOLOK," SOSTECH, vol. 1, 2021.

[19] F. Juniaty Simatupang, T. Wuryandari, M. Jurusan Statistika FSM Universitas Diponegoro, and S. Pengajar Jurusan Statistika, "KLASIFIKASI RUMAH LAYAK HUNI DI KABUPATEN BREBES DENGAN MENGGUNAKAN METODE LEARNING VECTOR QUANTIZATION DAN NAIVE BAYES," JURNAL GAUSSIAN, vol. 5, no. 1, pp. 99–111, 2016, [Online]. Available: http://ejournal-s1.undip.ac.id/index.php/gaussian