**MENENTUKAN PENERIMA BANTUAN RUMAH TIDAK LAYAK HUNI MENGGUNAKAN MACHINE LEARNING: *SYSTEMATIC LITERATURE REVIEW***

**LAPORAN PROGRAM INTERNSHIP II**

Laporan ini dibuat untuk memenuhi persyaratan kelulusan

matakuliah Program Internship II



**Dibuat oleh,**

**1.19.4.010 Deriska Fadilla Musdalifa**

**PROGRAM DIPLOMA IV TEKNIK INFORMATIKA**

**SEKOLAH VOKASI**

**UNIVERSITAS LOGISTIK BISNIS INTERNASIONAL  
BANDUNG  
2023**

***DETERMINING HOUSE ASSISTANCE RECIPIENTS USING MACHINE LEARNING: SYSTEMATIC LITERATURE REVIEW***

***INTERNSHIP PROGRAM REPORT II***

*This report was made to meet the graduation requirements*

*For the internship II program course*



***Created by,***

***1.19.4.010 Deriska Fadilla Musdalifa***

**DIPLOMA IV PROGRAM IN INFORMATICS ENGGINERING**

**SEKOLAH VOKASI**

**UNIVERSITAS LOGISTIK BISNIS INTERNASIONAL  
BANDUNG  
2023**

**LEMBAR PENGESAHAN**

**MENENTUKAN PENERIMA BANTUAN RUMAH TIDAK LAYAK HUNI MENGGUNAKAN MACHINE LEARNING: *SYSTEMATIC LITERATURE REVIEW***

Deriska Fadilla Musdalifa 1194010

Laporan Program Internship II ini telah diperiksa, disetujui dan disidangkan di Bandung, 2023

Oleh:

|  |  |
| --- | --- |
| Penguji Pendamping, | Penguji Utama, |
| Nama Penguji Pendamping  NIK: | Roni Andarsyah, S.T., M.Kom., SFPC  NIK: 115.88.193 |

|  |
| --- |
| Menyetujui,  Koordinator Program Internship 1 |
| Mohamad Nurkamal Fauzan, S.T.,M.T.,SFPC  NIK. 113.80.159 |

**SURAT PERNYATAAN**

**TIDAK MELAKUKAN PLAGIARISME**

Yang bertanda tangan dibawah ini:

Nama : Deriska Fadilla Musdalifa

NPM 1194010

Program Studi : D4 Teknik Informatika

Judul : Menentukan Penerima Bantuan Rumah Tidak Layak Huni Menggunakan Machine Learning: *Systematic Literature Review*

Menyatakan bahwa:

1. Program Internship I saya ini adalah asli dan belum pernah diajukan untuk memenuhi kelulusan matakuliah Internship pada Program Studi D4 Teknik Informatika baik di Universitas Logistik dan Bisnis Internasional maupun di Perguruan Tinggi lainnya.
2. Program Internship I ini adalah murni gagasan, rumusan dan penelitian saya sendiri tanpa bantuan pihak lain, kecuali arahan pembimbing.
3. Dalam Program Internship I ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan-penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya ini, serta sanksi lainnya sesuai dengan norma yang berlaku diperguruan tinggi lain.

|  |
| --- |
| Bandung, 2023 |
| Deriska Fadilla Musdalifa  1194010 |

# ABSTRAK

Rumah layak huni diartikan sebagai rumah dengan fasilitas untuk kebutuhan aktivitas masyarakat sehari-hari. Dari sekian banyaknya data penerima bantuan RTLH hasilnya tidak sesuai dengan yang ada, seperti warga yang berhak tidak mendapatkan bantuan sedangkan orang kaya mendapatkan bantuan yang disebabkan karena pemerintah masih menggunakan perhitungan manual. Oleh karena itu studi ini menekankan pada calon penerima bantuan rumah tidak layak huni menggunakan metode machine learning. Tujuan dari klasifikasi studi ini yaitu untuk memberikan hasil keputusan calon penerima bantuan layak diterima atau tidak berdasarkan data. Laporan ini membahas tentang tinjauan prediksi algoritma machine learning untuk paper yang diterbitkan dari 2018 hingga 2023 di perpustakaan digital menggunakan aplikasi publish or perish. Kesimpulannya, penulis menyatakan bahwa model *K-Nearest Neighbor* dapat menjadi alternatif yang tepat untuk prediksi rumah tidak layak huni. Namun, perlu dilakukan penelitian lebih lanjut untuk mengembangkan algoritma ini dan memperluas penerapannya pada topik penerimaan rtlh.

Kata kunci: RTLH, *Machine Learning*, *Systematic Literature Review,* *K Nearest Neighbor*

***ABSTRACT***

*A livable house is defined as a house with facilities for the needs of the community's daily activities. Of the many data on recipients of RTLH assistance, the results are not in accordance with what is available, such as residents who are entitled to not get assistance, while the rich get assistance because the government is still using manual calculations. Therefore this study emphasizes prospective recipients of uninhabitable housing assistance using the machine learning method. The purpose of the classification of this study is to provide a decision on whether a candidate for assistance is eligible to be accepted or not based on data. This report discusses the review of machine learning algorithm predictions for papers published from 2018 to 2023 in digital libraries using the publish or perish application. In conclusion, the authors state that the K-Nearest Neighbor model can be an appropriate alternative for rtlh predictions. However, it is necessary to carry out further research to develop this algorithm and expand its application to the topic of RTLH receipts.*

*Keywords: RTLH, Machine Learning, Systematic Literature Review, K Nearest Neighbor*

# KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penyusunan laporan Intership dengan lancar dan tepat pada waktunya. Laporan program Internship ini penulis susun sebagai salah satu syarat untuk memenuhi persyaratan kelulusan program Internship II.

Semoga laporan yang berjudul **“Menen­tukan Penerima Bantuan Rumah Tidak Layak Huni dengan menggunakan Machine Learning: Systematic Literature Review”** dapat menjadi laporan yang sempurna beserta dengan hasil peneliatan yang selanjutnya dan dapat bermanfaat bagi para pembacanya.

Kami menyadari dalam laporan ini masih banyak kekurangan, oleh karena itu perlu kritik dan saran untuk penyempurnaan laporan selanjutnya.

Bandung, Mei 2023

Penulis

Deriska Fadilla Musdalifa

# DAFTAR ISI

[ABSTRAK ii](#_Toc134747131)

[KATA PENGANTAR iv](#_Toc134747132)

[DAFTAR ISI v](#_Toc134747133)

[DAFTAR GAMBAR vi](#_Toc134747134)

[BAB I PENDAHULUAN 1](#_Toc134747135)

[1.1 Latar Belakang 1](#_Toc134747136)

[1.2 Identifikasi Masalah 2](#_Toc134747137)

[1.3 Tujuan dan Manfaat 2](#_Toc134747138)

[1.4 Ruang Lingkup 3](#_Toc134747139)

[1.5 Penelitian Sebelumnya 4](#_Toc134747140)

[1.6 Sistematika Penulisan 5](#_Toc134747141)

[BAB II LANDASAN TEORI 6](#_Toc134747142)

[2.1 State of The Art 6](#_Toc134747143)

[2.2 Rumah Tidak Layak Huni 10](#_Toc134747144)

[2.3 Model Prediksi 10](#_Toc134747145)

[2.4 Klasifikasi 11](#_Toc134747146)

[2.5 Machine Learning 11](#_Toc134747147)

[2.6 Evaluasi Model 11](#_Toc134747148)

[BAB III METODOLOGI PENELITIAN 13](#_Toc134747149)

[3.1 Diagram Alur Metodologi Penelitian 13](#_Toc134747150)

[3.2 Tahapan-Tahapan Diagram Alur Metodologi Penelitian 13](#_Toc134747151)

[3.2.1 Mengajukan Pertanyaan Penelitian 13](#_Toc134747152)

[3.2.2 Melakukan Pencarian Literatur 14](#_Toc134747153)

[3.2.3 Menentukan Kriteria Penelitian 14](#_Toc134747154)

[3.2.4 Melakukan Ekstraksi Data 15](#_Toc134747155)

[BAB IV PENUTUP 5](#_Toc134747156)

[4.1 Kesimpulan 5](#_Toc134747157)

[4.2 Saran 5](#_Toc134747158)

[DAFTAR PUSTAKA 6](#_Toc134747159)

# DAFTAR GAMBAR

[Gambar 3. 1 Tahap Perencanaan Tinjauan Pustaka 13](#_Toc134747091)

[Gambar 3. 2 Grafik Publikasi 15](#_Toc134747092)

# PENDAHULUAN

## Latar Belakang

Rumah tidak layak huni adalah sebuah kondisi di mana sebuah rumah atau tempat tinggal tidak memenuhi standar minimal kelayakan sebagai tempat tinggal manusia yang aman, sehat, dan layak huni. Rumah tidak layak huni bisa disebabkan oleh berbagai faktor seperti kekurangan fasilitas dasar, kerusakan struktural, sanitasi yang buruk, kebakaran, banjir, dan lain sebagainya. Di banyak negara di seluruh dunia, masalah rumah tidak layak huni merupakan masalah yang serius dan mempengaruhi kualitas hidup banyak orang. Masalah ini biasanya terjadi pada keluarga dengan tingkat pendapatan rendah yang tidak mampu memperbaiki atau membangun rumah mereka sendiri yang layak huni. Selain itu, faktor lingkungan seperti gempa bumi, banjir, atau bencana alam lainnya juga dapat membuat rumah menjadi tidak layak huni. Rumah tidak layak huni dapat memiliki dampak yang signifikan pada kesehatan dan kesejahteraan manusia. Kondisi yang tidak aman dan tidak sehat dapat menyebabkan infeksi, penyakit, dan bahkan kematian. Selain itu, ketidaknyamanan yang diakibatkan oleh rumah tidak layak huni dapat menyebabkan masalah psikologis seperti stres, depresi, dan gangguan mental lainnya. Oleh karena itu, penting bagi pemerintah, organisasi non-pemerintah, dan individu untuk bekerja sama untuk meningkatkan kondisi perumahan bagi masyarakat yang kurang mampu. Tindakan seperti penyediaan program perumahan yang terjangkau, perbaikan rumah yang sudah ada, dan pemberian bantuan dalam situasi darurat dapat membantu mengurangi jumlah rumah tidak layak huni dan meningkatkan kualitas hidup manusia.

Prediksi machine learning rumah tidak layak huni adalah penggunaan teknologi machine learning untuk memprediksi kondisi rumah yang mungkin tidak layak huni. Teknologi machine learning menggunakan algoritma yang dapat mempelajari pola dari data dan memprediksi hasil berdasarkan data historis. Dalam konteks rumah tidak layak huni, teknologi machine learning dapat membantu dalam memperkirakan kondisi perumahan yang mungkin memerlukan perbaikan atau renovasi. Terdapat beberapa penelitian mengenai penggunaan algoritma machine learning untuk memprediksi penerimaan bantuan rumah tidak layak huni. Dalam hal ini, diperlukan pengetahuan dengan cara mengkaji penelitian yang sudah ada dengan metode SLR (*Systematic Literatur Review*). Dengan menggunakan SLR, peneliti dapat mengetahui algoritma machine learning terbaik yang dapat diterapkan pada topik rumah tidak layak huni. Untuk mengetahui prediksi algoritma machine learning untuk yang terbaik, peneliti memilih *k-fold cross validation* serta evaluasi dan validasi hasil menggunakan confusion. Berdasarkan pernyataan di atas, peniliti melakukan literature review pada paper menggunakan metode SLR untuk mengetahui algoritma machine learning terbaik dan dapat diimplementasikan pada topik penerimaan bantuan rumah tidak layak huni.

## Identifikasi Masalah

Rumusan masalah pada penelitian ini dapat diajukan sebagai berikut:

Bagaimana cara mengetahui jenis *machine learning* yang digunakan untuk menentukan penerimaan bantuan rumah tidak layak huni?

Bagaimana mengetahui metrik evaluasi model *machine learning* yang paling banyak digunakan untuk diimplementasikan pada prediksi penerimaan bantuan rumah tidak layak huni?

## Tujuan dan Manfaat

Adapun tujuan pada penelitian ini sebagai berikut:

Mengetahui algoritma *machine learning* yang paling banyak digunakan untuk diimplementasikan pada prediksi penerimaan bantuan rumah tidak layak huni.

Mengetahui metrik evaluasi model *machine learning* yang paling banyak digunakan untuk diimplementasikan pada prediksi penerimaan bantuan rumah tidak layak huni.

Manfaat dari penelitian ini yaitu sebagai berikut:

*Literature Review* yang dibuat dapat digunakan untuk menentukan model yang akan diterapkan.

Dapat digunakan untuk menentukan metrik evaluasi model *machine learning* yang paling banyak digunakan.

Dengan adanya pemodelan *machine learning*, dapat membantu dinas sosial dalam memberikan keputusan masyarakat yang layak mendapatkan bantuan rumah tidak layak huni.

## Ruang Lingkup

Agar penelitian ini dapat dilakukan secara terarah dan tidak keluar dari pokok permasalahan, maka ruang lingkup yang digunakan yaitu sebagai berikut:

Sebagai syarat untuk memenuhi nilai matakuliah Internship II.

Paper yang digunakan harus sesuai dengan kata kunci “*Klasifikasi*”, “Prediksi”, “RTLH”, “*Machine Learning*” dan “bantuan sosial”.

## Penelitian Sebelumnya

Berikut merupakan beberapa penelitian yang telah dilakukan sebelumnya dengan topik yang sama yaitu:



## Sistematika Penulisan

Penyajian laporan program Internship ini dibagi menjadi beberapa bab yang memiliki pembahasannya masing-masing. Berikut adalah pemaparan dari setiap bab tersebut:

BAB I PENDAHULUAN. Pada bab ini akan dibahas mengenai latar belakang dibuatnya penelitian. Kemudian akan di identifikasi masalah apa saja yang berkaitan dengan penelitian berdasarkan latar belakang, lalu menentukan tujuan dibuatnya penelitian ini berdasarkan dari identifikasi masalah.

BAB II LANDASAN TEORI. Pada bab ini berisi uraian-uraian definisi yang digunakan dalam penelitian ini. Bersumber dari jurnal yang terkait dengan penyusunan proposal program Internship II.

BAB III METODOLOGI PENELITIAN. Pada bab ini berisi berisi mengenai diagram alur metodologi penelitian dan tahapan-tahapan diagram alur metodologi penelitian yang akan digunakan.

BAB IV PENUTUP. Pada bab ini berisi kesimpulan dari hasil penelitian dan saran yang bersifat membangun dari kesimpulan hasil yang diperoleh.

# LANDASAN TEORI

## State of The Art

Berdasarkan Studi Literatur, peneliti mengumpulkan 25 literatur yang berhubungan dengan judul penerima bantuan rumah tidak layak huni menggunakan data mining. Menurut [1] pada penelitiannya yang berjudul Penerapan data mining untuk mengklasifikasi penerima bantuan pkh desa wae jare menggunakan merode naïve bayes mempperoleh akurasi sebanyak 82,14% menggunakan teknik confusion matrix. Perhitungan tersebut menggunakan aplikasi rapid miner. Menurut [2] dalam penelitiannya yang berjudul Klasifikasi penerima bantuan program rehabilitasi rumah tidak layak huni menggunakan algoritme K-Nearest Neighbor. Data yang diperolah sebanyak 2962 dengan 22 atribut yang kemudian digunakan sebagai data latih dan uji sehingga mendapatkan akurasi tertinggi dengan 13 atribut data dengan variable yang ditentukan disperkim yang meliputi aspek keselamatan, struktur bangunan dan kesehatan dengan akurasi 97,93%, nilai k 5, presisi 96,88%, recall 99,63% dan AUC 0,964. Hasil klasifikasi terhadap data uji menggunakan matrik confussion. Menurut [3] dalam penelitiannya yang berjudul Implementasi Metode JST-Backpropagation Untuk Klasifikasi Rumah Layak Huni (Studi Kasus: Desa Kidal Kecamatan Tumpang Kabupaten Malang) menggunakan 160 data yang terbagi menjadi 2 yaitu layak dan tidak layak. Metode Backpropagation merupakan metode klasifikasi dengan kinerja yang sangat baik dan inisialisasi bobot awal menggunakan Nguyen Widrow dalam penelitian ini. Dengan pengujian menggunakan metode k fold cross validation mendapatkan akurasi tertinggi sebesar 57,14% dengan k 10. Pengujian akhir dari penelitian ini menghasilkan nilai rata-rata akurasi akurasi tertinggi sebesar 59% dengan menggunakan 15 input layer, 3 hidden layer,learning rate sebesar 0,2. Kemudian menurut [4] dengan judul papernya Penerapan Algoritma C4.5 Untuk Klasifikasi Penentuan Penerimaan Bantuan Langsung di Desa Ciomas mendapatkan presentase 74,9%. Aplikasi pendamping rapid miner digunakan dalam penelitian [5] untuk menguji akurasi sistem yang buat. Pengujian dilakukan dengan menyiapkan data training sebanyak 180 data dan data testing sebanyak 33 data yang dipilih secara random. Hasil pengujian akurasi yang diperoleh dengan metode naive bayesian cukup tinggi yaitu sebesar 93,94%, dan tingkat error sebesar 6,06%. Menurut [6] dalam penelitiannya berjudul Sistem Pendukung Keputusan Penerimaan Bantuan Bedah Rumah Menggunakan Metode Weight Product (WP) Dan Simple Additive Weighting (SAW) (Studi Kasus : Desa Semuli Raya Kecamatan Abung Semuli) memperoleh nikai akurasi 31,9% dengan metode WP dan 93,8% dengan SAW. Pada penelitian [7] dengan judul Penerapan Metode Data Mining C4.5 dalam Penentuan Kelayakan Rehabilitas Rumah Warga menghasilkan akurasi sebesar 83,33% menggunakan Teknik split validation dengan menggunakan bantuan software rapidminer. Menurut [8] pada penelitianya berjudul Terakreditasi SINTA Peringkat 4 Sistem Pendukung Keputusan Penerima Bantuan Rumah Layak Huni Menggunakan FMADM dan SAW mendapatkan akurasi sebesar 95,44% untuk metode SAW dann 94,24% untuk FMADM. Menurut [9] pada judul Algoritma K-Nearest Neighbor Untuk Klasifikasi Rumah Layak Atau Tidak Layak huni mendapatkan akurasi sebesar 96,25% dengan metode 10-fold cross validation mendapatkan presisi 100% dengan recall 100% dan k 13. Menurut [10] pada jurnal yang berjudul aplikasi klasifikasi enerima kartu Indonesia sehat menggunakan algoritma naïve bayes classifier dengan percobaan 13 kali mendapatkan akurasi 94,78% dengan presisi 98,86% dan recall 90,98%. Menurut [11] pada jurnal yang berjudul klasifikasi penerima dana bantuan desa menggunakan metode KNN mendapatkan akurasi sebesar 81,25 dengan menggunakan pengujian k fold validation dan menggunakan rapid miner untuk menerapkannya. Menurut [12] pada judul Implementasi Metode Support Vector Machine (SVM) Untuk Klasifikasi Rumah Layak Huni (Studi Kasus: Desa Kidal Kecamatan Tumpang Kabupaten Malang) dengan hasil pengujian menghasilkan rata-rata akurasi sebesar 98,75% dengan menggunakan metode pengujian K-fold Cross Validation dengan nilai k = 10, serta parameter metode SVM antara lain yaitu 𝜆 = 0,5, 𝛾 = 0,001, 𝐶 = 1, 𝑑 = 2, iterasi maksimum = 10 literasi dan menggunakan kernel Polynomial of degree. Menurut [13] pada judul Pendekatan Machine Learning dalam Memprediksi Keluarga Penerima Program PKH menggunakan beberapa metode seperti 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. Menurut [14] dengan judul paper penerapan algoritma C4.5 untuk klasifikasi mahasiswa penerima bantuan social covid 19 mendapatkan kesimpulan Dengan mengumpulkan 500 data mahasiswa Universitas Tarumanagara digunakan 400 mahasiswa sebagai data pelatihaan dan 100 mahasiswa sebagai data pengujian dan hasil dari lima kali percobaan pengujian data tersebut menunjukkan bahwa aplikasi ini sudah berfungsi dengan baik dalam mengklasifikasi mahasiswa calon penerima bantuan sosial dengan rata-rataaccuracy sebesar 89%, precision sebesar 90.16%, dan recall sebesar 83.27%. Menurut [15] pada judul jurnal implementasi algoritma random forest untuk menentukan penerima bantuan raskin. Hasil yang diperoleh menunjukkan algoritma klasifikasi RF memiliki nilai precision, recall, f-measure dengan nilai 97%, nilai accuracy sebesar 97,26% dan nilai ROC 0,970, lebih baik dari algoritma klasifikasi lainnya yaitu perbedaan sebesar 5,11% dengan algoritma klasifikasi support vector machine dan 8,87% dengan algoritma klasifikasi naive bayes. Akurasi sangat baik digunakan sebagai acuan kinerja algoritma apabila jumlah False Negative dan False Positive jumlah nya mendekati. Hasil penelitian ini dibuktikan secara akurat dan sistematis menggunakan Receiver Operating Characteristic (ROC). Menurut [16] pada judul Analisa pebandingan algoritma naïve bayes, KNN dan SVM untuk klasifikasi status kemiskinan di jawa barat menghasilkan nilai tertinggi pada algoritma naïve bayes sebesar 92,56%, KNN 92,36% dan SVM 90,4%. Menurut [17] pada judul algoritma SVM untuk klasifikai ekonomi penduduk penerima bantuan pemerintah di kecamatan simpang raya Sulawesi tengah. Penelitian ini menggunakan data yang terdiri dari data latih sebanyak 320 data dan data uji sebanyak 80 data. Berdasarkan hasil uji coba keakuratan algoritma SVM didapatkan hasil akurasi sebesar 98%. Kemudian menurut [18] pada jurnal perbandingan algoritma C4.5 dan Naïve bayes untuk menentukan kelayakan penerima bantuan program keluarga harapan. dua algoritma tersebut menggunakan tools RapidMiner didapatkan hasil yaitu algoritma C4.5 menghasilkan nilai akurasi sebesar 91,25% dan nilai AUC sebesar 0,930 dengan tingkat diagnosa Excellent Classification sedangkan algoritma Naïve Bayes menghasilkan nilai akurasi sebesar 87,11% dan nilai AUC sebesar 0,923 dengan tingkat diagnosa Excellent Classification. Sehingga didapat kesimpulan algoritma C4.5 merupakan algoritma yang baik untuk diterapkan pada kelayakan PKH. Berdasarkan judul [19] tentang Sistem Pendukung Keputusan Penerimaan Bantuan Sosial Program Keluarga Harapan (PKH) dengan Menggunakan Metode Naïve Bayes Classifier (Studi Kasus di Balai Desa Bendungan Kraton Pasuruan). Hasil penelitian menunjukkan bahwa dengan menggunakan metode naive bayesian classifier, dengan 50 data latih dan 10 data uji, tingkat akurasi mencapai 80%. Menurut [20] dalam makalah yang berjudul Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN. Hasil dari penelitian ini menggunakan metode Algoritma K-NN dengan 6.491 data penerima bantuan menghasilkan nilai accuracy 66,46% sedangkan nilai accurasy model algoritma Naïve Bayes sebesar 99,88%. Menurut [21] pada judul implementasi metode naïve bayes untuk klasifikasi penerima program keluarga harapan Dataset dalam penelitian ini adalah data penerima Program Keluarga Harapan tahun 2020 sebanyak 82 sampel. Hasil pemodelan klasifikasi dengan algoritma Naive Bayes menghasilkan nilai presisi untuk kelas positif 100%, untuk kelas negatif 77%, nilai recall untuk kelas positif 80%, untuk kelas negatif 100%, nilai f1-score untuk kelas positif 89%, untuk kelas negatif 87%, dan nilai akurasi 88%.

Berdasarkan literatur terpilih menujukkan bahwa pendekatan KNN dan NB merupakan model yang banyak digunakan dalam penelitian mengenai penerima bantuan rtlh. Pada literatur diatas menggunakan model k-fold cross validation dan confusion matrix serta untuk metrix evaluasi yang digunakan yaitu Akurasi, Presisi dan Recall.

## Rumah Tidak Layak Huni

Rumah yang layak huni merupakan kebutuhan dasar manusia, rumah yang layak huni dikatakan jika pemilik rumah memiliki rasa tentram dan nyaman tinggal di rumah tersebut. Rumah yang layak huni dimaknai sebagai rumah yang memiliki fasilitas untuk memenuhi kebutuhan manusia dalam melaksanakan aktivitas sehari-hari [22]. Permasalahan kemiskinan merupakan salah satu masalah yang sangat mendesak untuk ditangani, khususnya di wilayah yang sulit dijangkau oleh pemerintah. Salah satu ciri umum dari kondisi fisik masyarakat miskin adalah tidak memiliki akses prasarana dan sarana dasar lingkungan yang memadai, dengan kualitas perumahan dan pemukiman yang jauh di bawah kelayakan, serta mata pencaharian yang tidak menentu [23].

## Model Prediksi

Model prediksi adalah model matematika atau statistika yang dibuat untuk memprediksi nilai atau hasil tertentu berdasarkan data yang diberikan serta dapat digunakan untuk memprediksi berbagai macam hal, seperti hasil keuangan, kejadian masa depan, perilaku pelanggan, atau pergerakan harga pasar[24]

Contoh model prediksi termasuk model regresi, model klasifikasi, dan model clustering yang dimana model regresi digunakan untuk memprediksi nilai numerik, seperti harga rumah atau penghasilan seseorang, berdasarkan variabel yang lain, sedangkan model klasifikasi digunakan untuk memprediksi kategori atau label, seperti apakah sebuah email spam atau bukan, berdasarkan fitur-fitur yang ada, sementara model clustering digunakan untuk mengelompokkan data berdasarkan kemiripan mereka, tanpa adanya label yang sudah diketahui sebelumnya[25]

## Klasifikasi

Klasifikasi adalah salah satu topik utama dalam pembelajaran data mining atau machine learning. Klasifikasi adalah kumpulan data dimana data yang digunakan tersebut berisi kelas label atau target. Klasifikasi terdiri dari tiga tahap, yaitu pembangunan model, penerapan model dan evaluasi. Pembangunan model adalah membangun model menggunakan data latih yang telah memiliki atribut dan kelas. Kemudian, data tersebut diterapkan untuk mendefinisikan kelas atau objek data baru. Kemudian, data tersebut dievaluasi untuk melihat seberapa akurat model tersebut untuk mengembangkan dan menerapkan data baru tersebut. Proses klasifikasi terdiri dari dua fase, yaitu fase training dan fase testing[26].

## Machine Learning

Machine Learning adalah bidang studi yang berfokus pada perancangan dan analisis algoritma yang memungkinkan komputer mempelajari sekumpulan data kemudian membangun model berdasarkan data tersebut. Machine Learning juga dapat diartikan sebagai komputer yang memiliki kemampuan untuk belajar tanpa diprogram secara eksplisit. Program tersebut menggunakan data untuk membangun model dan membuat keputusan tentang model yang akan dibangun. Pembelajaran machine learning yang digunakan yaitu supervised learning. Supervised Learning adalah pembelajaran yang diarahkan/diawasi. Jika di analogikan pada proses pembelajaran, maka komputer atau mesin akan mempelajari data latih yang berisi label. Dan jika di analogikan melalui siswa dan guru, maka komputer ialah siswa yang belajar dan guru akan meminta siswa untuk belajar dari soal yang sudah memiliki solusi dan kunci jawaban. Pada intinya ketika melakukan proses pembelajaran, sistem diberikan data yang berisi solusi atau label yang akan diinginkan atau contoh kasus yang akan di selesaikan oleh sistem[27].

## Evaluasi Model

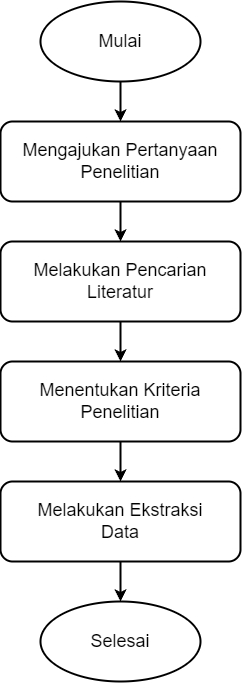
Evaluasi model adalah proses mengukur kinerja atau keefektifan model prediksi dengan menggunakan data yang sudah diketahui hasilnya untuk memastikan bahwa model yang dibangun dapat digunakan secara efektif untuk memprediksi nilai atau hasil tertentu pada data yang belum pernah dilihat sebelumnya[28]. Evaluasi model biasanya dilakukan dengan membagi data ke dalam dua kelompok yaitu data pelatihan dan data pengujian yang dimana data pelatihan digunakan untuk melatih model, sementara data pengujian digunakan untuk menguji kinerja model[29].

Kinerja model dapat dievaluasi dengan membandingkan hasil prediksi model dengan nilai sebenarnya dari data pengujian[30]. Evaluasi model yang baik dan akurat sangat penting dalam pengembangan model prediksi yang efektif untuk membantu memastikan bahwa model dapat digunakan untuk memprediksi hasil atau nilai yang akurat dan dapat diandalkan.

# METODOLOGI PENELITIAN

## Diagram Alur Metodologi Penelitian

Terdapat 4 tahapan dalam perencanaan tinjauan pustaka untuk mencari paper penelitian terkait seperti yang digambarkan pada Gambar 3.1



Gambar 3. 1 Tahap Perencanaan Tinjauan Pustaka

## Tahapan-Tahapan Diagram Alur Metodologi Penelitian

Adapun tahapan-tahapan dalam diagram alur metodologi yang digunakan terdapat beberapa tahap, berikut merupakan penjelasan setiap tahapanya:

### Mengajukan Pertanyaan Penelitian

Dalam tahap perencanaan, pertanyaan penelitian ditentukan untuk kontribusi penelitian dengan menjawab pertanyaan penelitian (PP):

* PP1: Apa saja algoritma *machine learning* untuk diterapkan pada topik bantuan rumah tidak layak huni?
* PP2: Apa saja metrik evaluasi model yang telah diterapkan untuk topik bantuan rumah tidak layak huni?

### Melakukan Pencarian Literatur

Setelah menentukan pertanyaan penelitian, kata kunci dan sumber data harus spesifik. Pertama, untuk mengidentifikasi kata kunci, pertanyaan penelitian dianalisis sebagai berikut: “RTLH”, “Prediction”, “Algorithm”, “Machine Learning” dan “bantuan sosial”. Selain itu dengan menambahkan “ATAU” dan “DAN” yang digunakan untuk mencari semua kemungkinan kombinasi dari kata kunci ini.

Pengumpulan data untuk kajian literatur artikel jurnal nasional yang diperoleh dari database *google scholar* menggunakan aplikasi *Harzing’s Publish or Perish Perish* (POP) yang digunakan untuk membantu peneliti dalam mencari, menganalisis, dan mengevaluasi jurnal, paper, serta publikasi ilmiah lainnya yang telah diterbitkan. Judul, abstrak, dan kata kunci dari semua artikel yang diindeks dapat dicari menggunakan istilah pencarian yang sudah ditentukan berdasarkan PP, dan pencarian dilakukan pada studi dari tahun 2018 hingga 2023.

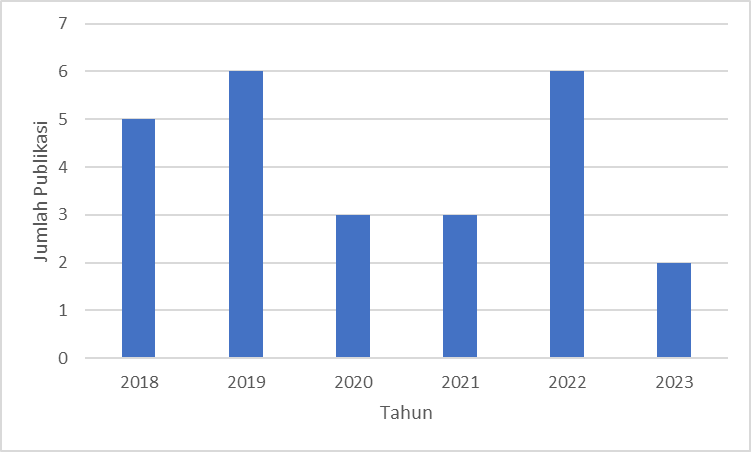
### Menentukan Kriteria Penelitian

Pada tahap ini dilakukan penyaringan penelitian untuk menilai secara menyeluruh relevansi dari kajian literatur ini adalah:

* Menyertakan artikel yang membahas tentang machine learning dan prediksi penerimaan bantuan rumah tidaklayak huni.
* Menyertakan jurnal dan buku.
* Menyertakan artikel yang diterbitkan dari 2018 hingga 2023.

### Melakukan Ekstraksi Data

21 referensi dipilih untuk ditinjau, hasil dari pencarian penelitian yaitu artikel tersebut diklarifikasikan berdasarkan jenis dan tahun penerbitan mulai tahun 2018 hingga 2023. Terlihat pada grafik 3.2 bahwa terjadi kenaikan pesat pada tahun 2022 dan penurunan pada tahun 2023.



Gambar 3. Grafik Publikasi

# PENUTUP

## Kesimpulan

Berdasarkan pencarian paper dengan kata kunci “*machine learning* penerimaan bantuan sosial” dengan batasan sebanyak 100 paper, ditemukan 25 paper yang telah memenuhi kriteria inklusi dan eksklusi. Berikut kesimpulannya:

Metode yang paling banyak digunakan pada area penelitian ini adalah *KNN (K-Nearest Neighbor),* dan *Naïve Baiyes*.

Evaluasi yang banyak digunakan yaitu akurasi, presisi dan recall.

## Saran

Saran untuk penelitian ini adalah sebagai berikut:

1. Perlu dilakukan pengkajian *paper* secara berkala agar hasil yang didapatkan baik.
2. Diperlukan penelitian lebih lanjut untuk membandingkan beberapa metode dengan data lebih dari 1000 data dengan data yang berbeda.

# DAFTAR PUSTAKA

[1] A. Ifon Purnama, A. Aziz, A. Sartika Wiguna, and K. Kunci, “PENERAPAN DATA MINING UNTUK MENGKLASIFIKASI PENERIMA BANTUAN PKH DESA WAE JARE MENGGUNAKAN METODE NAÏVE BAYES,” Oct. 2020. [Online]. Available: https://jurnal.machung.ac.id/index.php/kurawal

[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] R. Rizqiana, P. Putri, M. T. Furqon, and B. Rahayudi, “Implementasi Metode JST-Backpropagation Untuk Klasifikasi Rumah Layak Huni (Studi Kasus: Desa Kidal Kecamatan Tumpang Kabupaten Malang),” 2018. [Online]. Available: http://j-ptiik.ub.ac.id

[4] 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.

[5] N. A. Widiastuti, A. K. Zyen, and N. Safik, “PREDIKSI PENENTUAN PEMOHON KREDIT SEPEDA MOTOR MENGGUNAKAN ALGORITMA NAIVE BAYES,” *DISPROTEK*, vol. 10, 2019.

[6] F. Susanto *et al.*, “Sistem Pendukung Keputusan Penerimaan Bantuan Bedah Rumah Menggunakan Metode Weight Product (WP) Dan Simple Additive Weighting (SAW) (Studi Kasus : Desa Semuli Raya Kecamatan Abung Semuli),” *JTKSI*, vol. 03, 2020.

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

[8] S. K. Dirjen *et al.*, “Terakreditasi SINTA Peringkat 4 Sistem Pendukung Keputusan Penerima Bantuan Rumah Layak Huni Menggunakan FMADM dan SAW,” 2018.

[9] 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.

[10] A. A. Rahman and Y. I. Kurniawan, “APLIKASI KLASIFIKASI PENERIMA KARTU INDONESIA SEHAT MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER.”

[11] R. L. Hasanah *et al.*, “KLASIFIKASI PENERIMA DANA BANTUAN DESA MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR),” *Jurnal TECHNO Nusa Mandiri*, vol. 16, no. 1, p. 1, 2019, [Online]. Available: http://nusamandiri.ac.id/

[12] 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

[13] I. A. Sobari and R. A. Zuama, “Pendekatan Machine Learning dalam Memprediksi Keluarga Penerima Program PKH,” *Jurnal Teknik KomputerAMIK BSI*, vol. 9, 2023, doi: 10.31294/jtk.v4i2.

[14] E. E. Barito, ) Jap, T. Beng, and D. Arisandi, “Jurnal Ilmu Komputer dan Sistem Informasi.”

[15] I. Kurniawan, D. Cahya Putri Buani, W. Apriliah, R. Amegia Saputra, and P. Korespondensi, “IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK MENENTUKAN PENERIMA BANTUAN RASKIN IMPLEMENTATION OF RANDOM FOREST ALGORITHM FOR DETERMINING RECIPIENTS OF RASKIN,” vol. 10, no. 2, pp. 421–428, 2023, doi: 10.25126/jtiik.202396225.

[16] Akbar Anung, “ANALISA PERBANDINGAN ALGORITMA NAÏVE BAYES, K-NEAREST,” 2022. [Online]. Available: https://lib.mercubuana.ac.id/

[17] A. A. Kasim and M. Sudarsono, “Algoritma Support Vector Machine (SVM) untuk Klasifikasi Ekonomi PendudukPenerima Bantuan Pemerintah di Kecamatan Simpang Raya Sulawesi Tengah,” *Seminar Nasional APTIKOM*, 2019.

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

[19] 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.

[20] H. Putri, A. I. Purnamasari, A. R. Dikananda, O. Nurdiawan, and S. Anwar, “Penerima Manfaat Bantuan Non Tunai Kartu Keluarga Sejahtera Menggunakan Metode NAÏVE BAYES dan KNN,” *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 3, pp. 331–337, Dec. 2021, doi: 10.47065/bits.v3i3.1093.

[21] A. Asri *et al.*, “IMPLEMENTASI METODE NAIVE BAYES UNTUK KLASIFIKASI PENERIMA PROGRAM KELUARGA HARAPAN,” 2022.

[22] “rtlh1”.

[23] N. Kahar Teknik Informatika and S. Nurdin Hamzah Jambi Jln Kolonel Abunjani, “PENERAPAN METODE FUZZY MULTICRITERIA DECISION MAKING UNTUK SELEKSI PENERIMA BANTUAN RUMAH LAYAK HUNI (STUDI KASUS DI DESA SINGKAWANG JAMBI).”

[24] J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, “A comparison of random forest variable selection methods for classification prediction modeling,” *Expert Systems with Applications*, vol. 134. Elsevier Ltd, pp. 93–101, Nov. 15, 2019. doi: 10.1016/j.eswa.2019.05.028.

[25] A.K and E.G, “A review of data-driven building energy consumption prediction studies,” *Renewable and Sustainable Energy Reviews*, vol. 81, 2018.

[26] K.H and N.A.D, “Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma KNN,” *Computer Engineering, Science and System Journal*, p. 78, 2019.

[27] D.I, “Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python,” 2021.

[28] R.S, “Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning,” *Arxiv Journal*, 2018.

[29] R.F, “COVID-19 Future Forecasting Using Supervised Machine Learning Models,” 2018.

[30] M.A, “State of the art of machine learning models in energy systems, a systematic review," Energies,” *Energies (Basel)*, vol. 12, 2019.