

Pengaruh Pendidikan Terhadap Tingkat Kesehatan Masyarakat Kecamatan Suralaga Kabupaten Lombok Timur Menggunakan Algoritma Random Forest

Yahya^{1*}, Nurhidayati², L. M. Samsu³

^{1,2,3}Program Studi Sistem Informasi, Universitas Hamzanwadi

*ayhay7078@gmail.com

SDGs merupakan global yang menjadi perhatian masyarakat Internasional, tidak terkecuali Bangsa Indonesia. Banyak sekali kegiatan yang dilakukan oleh masyarakat Internasional terhadap penanganan masalah isu SDGs ini. Mulai dari tindakan nyata di lapangan sampai proses pendataan yang berkesinambungan. Di Indonesia, telah dilakukan pendataan secara berulang-ulang, namun tindak lanjut terhadap pendataan itu masing merasa kurang, oleh sebab itu melalui Universitas Hamzanwadi dengan menggunakan aplikasi KKN Tematik Desa Gemilang, juga telah berperan aktif dalam proses pendataan tersebut. Namun sampai saat ini, informasi tentang hasil pendataan dan tindak lanjutnya tidak terpapar dengan baik bahkan belum ada satu pun hasil pendataan yang telah diolah untuk memperoleh informasi yang jelas dan terukur. Melalui penelitian internal perguruan tinggi yang dilaksanakan oleh Universitas Hamzanwadi, dicoba untuk melakukan penelitian yang dimulai dengan proses pengolahan data SDGs yaitu pendidikan dan kesehatan untuk memberikan ilmu pengetahuan dan informasi tentang ada atau tidaknya pengaruh pendidikan terhadap kesehatan masyarakat Kecamatan Suralaga Kabupaten Lombok Timur, sehingga apabila pengaruh pendidikan tersebut mempunyai andil yang cukup besar, maka Pemerintah Kecamatan Suralaga dapat mengambil langkah-langkah yang kongkrit untuk perbaikan di masa mendatang. Dari hasil pengolahan data dapat disimpulkan bahwa pengaruh pendidikan terhadap tingkat kesehatan masyarakat sangat tinggi, menggunakan 90% data training dan 10% data testing dapat mencapai akurasi 88,61%, yang berarti 700 data menunjukkan trend positif pendidikan terhadap tingkat kesehatan masyarakat.

Kata kunci : SDGs, Kesehatan Masyarakat, Random Forest

Abstract

SDGs are a global concern of the international community, including the Indonesian nation. There are many activities carried out by the international community to address this SDGs issue. Starting from real actions in the field to a continuous data collection process. In Indonesia, data collection has been carried out repeatedly, but the follow-up to each data collection is felt to be lacking, therefore Hamzanwadi University, using the Gemilang Village Thematic KKN application, has also played an active role in the data collection process. However, up to now, information about data collection results and follow-up actions has not been provided properly, in fact none of the data collection results have been processed to obtain clear and measurable information. Through internal higher education research carried out by Hamzanwadi University, an attempt was made to carry out research starting with the SDGs data processing process, namely education and health, to provide knowledge and information about whether or not there is an influence of education on the health of the people of Suralaga District, East Lombok Regency, so that if the influence of education plays a significant role, the Suralaga District Government can take concrete steps to improve it in the future. From the results of data processing it can be concluded that the influence of education on the level of public health is very high, using 90% training data and 10% testing data can achieve an accuracy of 88.61%, which means 700 data show a positive trend of education on the level of public health.

Keywords : SDGs, Public Health, Random Forest

1. Pendahuluan

Masyarakat Suralaga memiliki karakteristik hidup yang berbeda-beda antara yang satu dengan

yang lainnya, hal ini karena pola kehidupannya yang terbentuk dari kehidupan di lapangan yang pada dasarnya mempunyai mata pencaharian

yang sangat kompleks, seperti : pedagang, petani, pegawai negeri sipil (PNS), buruh tani, buruh pasar, dan bahkan banyak yang menjadi Tenaga Kerja Indonesia (TKI) di beberapa Negara antara lain : Malaysia, Singapura, Brunei, Abu Dhaby, dan lain sebagainya.

SDGs merupakan global yang menjadi perhatian masyarakat Internasional, tidak terkecuali Bangsa Indonesia. Banyak sekali kegiatan yang dilakukan oleh masyarakat Internasional terhadap penanganan masalah isu SDGs ini. Mulai dari tindakan nyata di lapangan sampai proses pendataan yang berkesinambungan. Di Indonesia, telah dilakukan pendataan secara berulang-ulang, namun tindak lanjut terhadap pendataan itu masing merasa kurang, oleh sebab itu melalui Universitas Hamzanwadi dengan menggunakan aplikasi KKN Tematik Desa Gemilang, juga telah berperan aktif dalam proses pendataan tersebut. Namun sampai saat ini, informasi tentang hasil pendataan dan tindak lanjutnya tidak terpapar dengan baik bahkan belum ada satu pun hasil pendataan yang telah diolah untuk memperoleh informasi yang jelas dan terukur. Melalui penelitian internal perguruan tinggi yang dilaksanakan oleh Universitas Hamzanwadi, dicoba untuk melakukan penelitian yang dimulai dengan proses pengolahan data SDGs yaitu pendidikan dan kesehatan untuk memberikan ilmu pengetahuan dan informasi tentang ada atau tidaknya pengaruh pendidikan

terhadap kesehatan masyarakat Kecamatan Suralaga Kabupaten Lombok Timur, sehingga apabila pengaruh pendidikan tersebut mempunyai andil yang cukup besar, maka Pemerintah Kecamatan Suralaga dapat mengambil langkah-langkah yang kongkrit untuk perbaikan di masa mendatang.

Untuk membantu dan mempermudah dalam proses penelitian, maka telah berhasil dilakukan pendataan SDGs dengan 6 komponen yaitu ekonomi, kesehatan, kemiskinan, gender, lingkungan dan pendidikan. Dari ke 6 komponen yang adalah, maka peneliti akan menggabungkan dua komponen yaitu pendidikan dan kesehatan untuk memperoleh ilmu pengetahuan dan informasi yang dapat dipertanggung-jawabkan.

Data yang digunakan untuk penelitian kali ini bersifat private data, yang diperoleh dari aplikasi KKN Tematik Desa Gemilang. Yang sengaja dibuat untuk mendapatkan data yang akurat dan terintegrasi dengan baik. Pengambilan data menggunakan pengisian angket dan diberikan kepada seluruh masyarakat di wilayah Kecamatan Suralaga Kabupaten Lombok Timur. Untuk membantu mengklasifikasikan tingkat kesejahteraan masyarakat Suralaga di wilayah Lombok Timur, maka data yang dimiliki akan diklasifikasikan lalu diolah dengan algoritma Random Forest. Algoritma Random Forest memiliki konsep pohon keputusan dimana pada pohon keputusan tersebut memiliki prediksi kelas

dan output akhir dari sebuah keputusan dengan hasil yang maksimum^[1]

2. Tinjauan Pustaka

2.1. Penelitian Terkait

Dalam menjalankan penelitian ini, penulis merujuk pada penelitian terdahulu yang berkaitan dengan topik penelitian ini antar lain :

- Penelitian yang dilakukan oleh Bahtiar Imran dkk, dalam memprediksi dan mengklasifikasikan kanker payudara jinak dan ganas menggunakan tiga metode klasifikasi yaitu Random Forest, Naïve Bayes dan Ada Boost. Hasil prediksi dari penelitian tersebut menempatkan algoritma Random Forest dalam urutan pertama dengan hasil akurasi 100%. Penelitian ini digunakan untuk memprediksi dan mengklasifikasikan penyakit kanker payudara antara yang jinak dan ganas, selain itu juga digunakan untuk mendeteksi resiko penyakit kanker payudara agar pasien mendapatkan tindakan pencegahan di awal^[2]
- Sedangkan Veronica Retno Sari dkk melakukan penelitian dengan membandingkan prediksi kualitas kopi arabika dengan menggunakan algoritma SGD, Random Forest dan Naïve Bayes. Dari perbandingan algoritma tersebut SGD memiliki hasil akurasi tertinggi dengan 98% dan ketika menggunakan k-fold validation dan

AUC, hasil akurasi menjadi 99% ^[3] Dari penelitian yang dilakukan membuat peneliti mencoba menggunakan algoritma Random Forest dan Naïve Bayes dalam dataset yang ada. Apakah dua algoritma ini dapat menghasilkan parameter yang lebih baik dari penelitian sebelumnya. Karena tidak semua algoritma pada data mining memberikan performance model yang baik dalam mengklasifikasikan dataset.

- Penelitian selanjutnya dilakukan oleh Mulkan Azhari dkk, dengan melakukan perbandingan akurasi, recall dan presisi kalsifikasi pada algoritma C4.5, Random Forest, SVM, dan Naïve Bayes. Pada penelitian ini algoritma random Forest memiliki hasil akurasi terkecil dibandingkan tiga algoritma lainnya. Hal ini yang membuat peneliti ingin melakukan perbandingan hasil dari performance algoritma Random Forest dan Naïve Bayes menggunakan dataset yang dimiliki. Penelitian sebelumnya dilakukan dengan menggunakan 200 data dengan perbandingan 70% data training dan 30% data testing, dan penelitian yang akan dilakukan memiliki data lebih banyak dari ini. Pada penelitian sebelumnya hasil akurasi yang didapat 95% untuk SVM, C4.5 dan Naïve Bayes memiliki akurasi yang sama yaitu 86,67% dan yang terakhir adalah random Forest dengan nilai akurasi 83,33% ^[4]

- Berikutnya Komparasi Algoritma Random Forest, Naïve Bayes dan K-NN dalam mengklasifikasikan data Penyakit Jantung, penelitian ini dilakukan oleh Amril Samosir dkk dengan menggunakan 304 dataset penyakit jantung. Pengolahan data pada penelitian dengan jumlah yang cukup besar dapat dilakukan menggunakan model tertentu sehingga dapat menghasilkan nilai yang akurat dan cepat. Beberapa algoritma digunakan agar dapat membandingkan hasil dari kinerja setiap algoritma yang digunakan terutama dalam mengklasifikasi kan data. Dari penelitian ini hasil tertinggi dimiliki oleh naïve bayes yaitu 0,91 AUC, 0,84 CA, 0,84 F1, 0,839 Precision dan 0,84 Recall [5]
- Dan penelitian selanjutnya adalah prediksi penerimaan beasiswa dengan menggunakan teknik Data Mining di Universitas Muhammadiyah Pringsewu oleh Baskoro dkk. Data yang digunakan dalam penelitian ini kurang lebih sebanyak 3500 data mahasiswa. Penelitian dilakukan karena pemberian beasiswa yang tidak tepat sasaran, sehingga menyebabkan dana beasiswa kurang maksimal. Selain itu juga proses pengolahan data masih dilakukan secara manual, oleh sebab itu dilakukan pengolahan data yang lebih efisien dan akurat. Dari hasil penelitian yang dilakukan nilai akurasi tertinggi dimiliki oleh algoritma Random Forest yaitu sebesar 94,23%^[6] Penelitian ini menjadi acuan peneliti untuk melakukan pengolahan data menggunakan algoritma yang sama yaitu Random Forest, tetapi untuk mendapatkan performance yang lebih baik peneliti menggunakan dua algoritma agar dapat melakukan perbandingan hasil.
- Berikutnya penelitian yang dilakukan oleh V. Jackins dkk, menggunakan algoritma Random Forest Classifier dan Naïve Bayes yang digunakan untuk mengklasifikasikan beberapa penyakit seperti diabetes, jantung dan kanker. Data tersebut digunakan memprediksi atau mendiagnosis kondisi Kesehatan pasien terhadap penyakit yang telah diklasifikasikan, dengan hasil akhir algoritma Random Forest memiliki nilai akurasi tertinggi dari yang lainnya^[7]
- Penelitian yang dilakukan oleh Ramdani dkk bertujuan untuk membandingkan metode Naïve Bayes dan Random Forest dalam menentukan prestasi belajar siswa pada jurusan RPL di SMK Swasta Siti Banun Sigambal, karena prestasi belajar merupakan salah satu aspek terpenting dalam bidang Pendidikan. Demi melakukan pembelajaran yang bermutu di sekolah maka dilakukan analisis dan kinerja siswa dalam belajar dengan menggunakan teknik data mining diantaranya menggunakan dua algoritma naïve bayes dan random forest sebagai

- perbandingan hasil nantinya. Untuk memprediksi data error yang dimiliki maka digunakan k-Fold Validation, dan membagi antara data training dan data testing, sehingga menghasilkan performance tertinggi dimiliki oleh Random Forest dengan hasil akurasi 95,12%^[8]
- Untuk penelitian selanjutnya dilakukan oleh Hiya Nalatissifa dkk, dengan melakukan perbandingan kinerja algoritma klasifikasi Naïve Bayes, Support Vector Machine (SVM), dan Random Forest untuk memprediksi ketidakhadiran di tempat kerja. Ketidakhadiran merupakan permasalahan an bagi sebuah perusahaan karena diartikan penundaan tugas bagi karyawan. Banyak faktor yang menyebabkan terjadinya ketidakhadiran. Hal ini yang mendorong untuk dijadikan penelitian. Dataset yang digunakan diperoleh dari UCI Machine Learning. Hasil dari penelitian mendapatkan algoritma Random Forest menduduki nilai tertinggi untuk akurasi 99,38%, presisi 99,42% dan recall 99,39%^[9]
 - Komparasi akurasi pada Naïve Bayes dan Random Forest dalam klasifikasi penyakit liver pernah dilakukan oleh Ahmadi Firmansyah dkk. Penelitian tersebut dilakukan untuk mengidentifikasi penyakit liver melalui diagnosis sebagai proses untuk menentukan sifat yang dapat membedakannya dari keadaan yang terjadi. Data diambil dari Indian Liver Patient Dataset yang diperoleh dari UCI Machine Learning Repository dan memiliki 583 record data. Hasil dari pengujian didapat algoritma Random Forest memiliki nilai akurasi terbaik yaitu sebesar 70,60%, sedangkan Naïve Bayes memperoleh akurasi sebesar 55,80%^[10]
 - Penelitian selanjutnya dilakukan oleh Ivana Alhabib dkk, dengan melakukan komparasi metode Deep Learning, Naïve Bayes dan Random Forest dalam memprediksi penyakit jantung. Penelitian ini dilakukan untuk melakukan pemodelan dan memecahkan masalah dalam memprediksi terjadinya serangan jantung pada pasien. Data penelitian diambil dari Kaggle, dari data yang diolah menggunakan tiga algoritma, Deep Learning memiliki tingkat akurasi tertinggi sebesar 83,49%^[11]
 - Model klasifikasi data juga dilakukan pada penelitian Klasifikasi penerima program Indonesia Pintar menggunakan algoritma Naïve Bayes dan Random Forest yang dilakukan oleh Andi Suandi dkk. Penelitian dilakukan bertujuan untuk melakukan klasifikasi penerima Program Indonesia Pintar agar tepat sasaran. Algoritma yang digunakan dalam penelitian ini adalah Naïve Bayes dan Random Forest. Dataset yang

digunakan menggunakan enam atribut, hasil akurasi yang didapat 99,96% dari algoritma Naïve Bayes dan 78,42% dari hasil algoritma Random Forest. Sehingga dapat disimpulkan performance terbaik adalah algoritma Naïve Bayes^[12]

- Penelitian selanjutnya adalah perbandingan Klasifikasi SMS Berbasis Support Vector Machine, Naïve Bayes Classifier, Random Forest dan Bagging Classifier yang dilakukan oleh Devi Irawan dkk. Penelitian dilakukan bertujuan untuk menyelesaikan masalah yang sering terjadi pada SMS yang teridentifikasi spam, selain itu juga penelitian dilakukan untuk memisahkan dan menyaring SMS spam dan Non spam. Dengan menggunakan empat algoritma maka diperoleh hasil akurasi tertinggi adalah 97,4% dimiliki oleh algoritma Bagging Classifier^[13]
- Kesehatan masyarakat merupakan bagian dari pola hidup yang digunakan untuk mencegah penyakit, memperpanjang masa hidup, dan meningkatkan kualitas sumber daya manusia. Oleh sebab itu, kesehatan masyarakat ditinjau dari berbagai aspek sangat perlu untuk dikaji, agar tujuan dan sasaran yang hendak dicapai dapat terlaksana. Salah satu kajian yang dikemukakan adalah penggunaan alat kontrasepsi bagi masyarakat pedesaan,

khususnya masyarakat pedesaan yang berada di Kecamatan Suralaga Kabupaten Lombok Timur – Nusa Tenggara Barat. Penggunaan alat kontrasepsi yang merupakan salah satu cara yang digunakan dalam penerapan Keluarga Berencana masih berada pada tingkat yang tergolong masih rendah. Berdasarkan data yang dimiliki yang berkaitan dengan hal tersebut, diperlukan adanya suatu analisis dan pengolahan data dalam menentukan keputusan yang akan diambil untuk mengetahui pengaruh alat kontrasepsi dalam peningkatan kesehatan masyarakat. Pengolahan dan analisis data dilakukan menggunakan algoritma Random Forest, yang disebabkan oleh karakteristik datasets yang digunakan sesuai. Untuk mendapatkan hasil atau performa yang terbaik dalam mengetahui tingkat akurasi, dari datasets yang digunakan memiliki 10 atribut yaitu : nama lengkap, umur, pendidikan terakhir, pasangan usia subur, usia hamil pertama, usia pertama melahirkan, siap kb, alat kontrasepsi, mengikuti program kontak, dan jarak melahirkan. Akurasi yang diperoleh sebanyak 71,99% dengan output pohon keputusan bahwa alat kontrasepsi yang paling banyak dan efisien untuk digunakan di wilayah Kecamatan Suralaga adalah Pil KB dan Suntik KB^[14].

2.2. Landasan Teori

Pada penelitian kali ini peneliti menggunakan metode klasifikasi yang digunakan untuk memprediksi tingkat kesehatan masyarakat Kecamatan Suralaga Kabupaten Lombok Timur?. Metode klasifikasi sendiri adalah bagian dari Data Mining. Sebuah proses pengolahan data mentah menjadi sebuah informasi yang bermanfaat dapat kita sebut Data Mining^[15] Data mining memiliki teknik dan sifat diantaranya :

- Classification (Predictive)
- Clustering (Descriptive)
- AssociationRule Discovery (Descriptive)
- Regression (Predictive)
- DeviationDetection (Predictive) (Eskal, n.d.)

Sedangkan klasifikasi adalah sebuah proses analisa yang dapat menciptakan model-model yang dapat mendeskripsikan kelas yang ada pada data dan merupakan salah satu teknik dari data mining yang biasa digunakan dalam pembelajaran^[16]

Mengimplementasikan dataset yang menjadi data training pada model yang ditentukan dan juga sudah dilakukan pengujian testing data adalah sebuah proses kalsifikasi^[17]

Metode klasifikasi yang digunakan adalah untuk memprediksi data set. Prediksi dan klasifikasi hampir memiliki kesamaan, yang membuat berbeda adalah prediksi memiliki nilai dari hasil yang akan datang^[18]

Metode klasifikasi yang digunakan menggunakan algoritma Random forest, dimana algoritma ini adalah sebuah algoritma machine learning yang digunakan untuk pengambilan keputusan yang berbentuk gabungan dari pohon keputusan dengan cara seleksi, dimana kumpulan pohon keputusan dilatih menggunakan metode Bagging untuk mendapatkan hasil dari keseluruhan^[19]

Dalam algoritma machine learning seperti Random Forest output yang dihasilkan dari metode klasifikasi dibagi menjadi empat yaitu : *True Positive (TP)*, *True Negative (TN)*, *False Negative (FN)*, dan yang terakhir adalah *False Positive (FP)*. Dalam machine learning tingkat kesalahan dari pengklasifikasian diberi simbol E yang merupakan frekuensi dari tingkat *error*. Tingkat kesalahan tersebut dapat dihitung menggunakan persamaan (1).

$$E = \frac{FP+FN}{FP+FN+TP+TN} \dots \dots \dots (1)$$

Selain tingkat kesalahan, hasil akurasi yang merupakan frekuensi dari klasifikasi dataset juga dapat dihitung menggunakan persamaan (2).

$$E = \frac{TP+TN}{FP+FN+TP+TN} \dots \dots \dots (2)$$

Dan persamaan yang ke (3) adalah presisi yang merupakan hasil persentase dari *True Positif* dari semua dataset yang mendapatkan label positif

$$Pr = \frac{TP}{FP+TP} \dots \dots \dots (3)$$

Proses di atas adalah proses evaluasi pada algoritma random forest^[20]

Selain Random Forest algoritma yang kedua adalah Naïve Bayes. Algoritma Naïve Bayes adalah sebuah teknik machine learning yang didasari oleh teori probabilitas yang berhubungan dengan probabilitas sebelumnya, sehingga dapat memberikan keputusan secara matematis dan menciptakan sebuah model yang sederhana dan efisien^[21]

Nilai probabilitas dalam teorema Bayes seperti pada persamaan 4 :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \dots\dots\dots (4)$$

Dimana :

P(H) : nilai probabilitas dari prior dari hipotesis sampel

P(X) : evidence dari probabilitas data latih

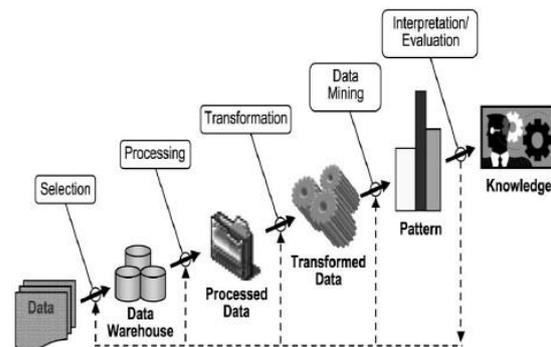
P(H|X) : Nilai probabilitas H yang mempengaruhi X

P(X|H) : probabilitas X kepada H atau dapat disebut dengan likelihood.

Dalam metode klasifikasi ada sebuah metode yang digunakan untuk mengukur kinerja metode klasifikasi itu yaitu Confusion matrix. Dimana Confusion matrix yang mengandung informasi digunakan untuk membandingkan hasil klasifikasi yang dikerjakan oleh sistem dengan hasil yang sesungguhnya^[22]

2.3. Tahapan Penelitian

Data Mining merupakan salah satu dari rangkaian *knowledge discovery in database (KDD)*. Adapun serangkaian proses tahapan dalam Data Mining tersebut sebagai berikut:



Gambar 1. Tahapan data mining

a) Pembersihan Data (*Data Cleaning*)

Merupakan proses untuk membuang data yang tidak konsisten dan menghilangkan *noise*. Data yang diperoleh baik dari database suatu perusahaan maupun hasil eksperimen memiliki isian yang tidak sempurna seperti data yang hilang, data tidak valid atau bahkan sekedar salah ketik. Data-data yang tidak relevan lebih baik dibuang, pembersihan data sangat penting dilakukan karena akan mempengaruhi performansi dari teknik data mining.

b) Integrasi Data (*Data Integration*)

Merupakan penggabungan data dari berbagai database ke dalam satu database baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu database tetapi juga berasal dari beberapa database

atau file teks. Integrasi dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan, dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya.

c) Seleksi Data (*Data Selection*)

Data yang ada pada database seringkali tidak semuanya dipakai. Oleh karena itu, hanya data yang sesuai untuk dianalisis yang akan diambil dari database.

d) Transformasi Data (*Data Transformasi*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan.

e) Proses Mining

Proses mining merupakan proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

f) Evaluasi Pola (*Pattern Evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi di evaluasi untuk menilai apakah

hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternative yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining yang lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil diluar dugaan yang mungkin bermanfaat.

3. Metode Penelitian

Dalam tahap pengumpulan data, Penelitian ini menggunakan metode pengumpulan data dua tipe yaitu pengumpulan data primer dan data sekunder. Data primer adalah data yang dikumpulkan pertama kali dan untuk melihat apa yang sesungguhnya terjadi. Data Sekunder adalah data yang sebelumnya pernah dibuat oleh seseorang baik diterbitkan atau tidak. dalam pengumpulan data primer, peneliti menggunakan metode observasi dan interview, dengan menggunakan data yang berhubungan dengan penggunaan data Kesehatan Masyarakat secara langsung kepada pihak yang terlibat secara langsung. sedangkan dalam pengumpulan data sekunder menggunakan jurnal,buku, publikasi, dan lain-lain.

Sumber data yang dikumpulkan langsung oleh pembuat laporan disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder. data yang diperoleh adalah

data sekunder karena diperoleh melalui proses pendataan pada pelaksanaan KKN Tematik Desa Gemilang Universitas Hamzanwadi pada tahun 2019. Data yang dikumpulkan adalah data Kesehatan Kepala keluarga pada tahun 2019 data yang terkumpul sebanyak 790 data

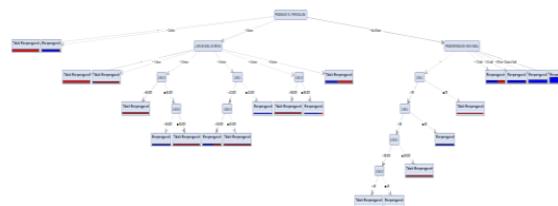
4. Hasil dan Pembahasan

Pada bagian ini, dicantumkan atau dilaporkan tentang hasil yang telah bisa dicapai dalam progres penelitian internal ini. Progres yang telah dicapai adalah telah berhasil mengumpulkan data masyarakat Kecamatan Suralaga sebagaimana ditunjukkan pada tabel 1 di bawah ini.

Tabel 1. sampel dataset

No	Nama	Umur	Pend. Terakhir	Putus Sekolah
1	Supardi	55	SMP	Ya
2	Muh.taufik Nasrullah	47	SMP	Ya
3	SIDIN	69	SD	Tidak
4	Bahrul Fahmi	41	S1	Tidak
5	Hasyim	42	SD	Ya

Hasil yang diperoleh dengan melakukan percobaan yang berbeda-beda adalah nilai akurasi yang berbeda bahkan telah menunjukkan bahwa akurasi tertinggi diperoleh adalah 88,61% yang dapat digambarkan seperti pada pohon keputusan (Random forest) di bawah ini.



Gambar 2. Pohon keputusan

5. Kesimpulan

Dari hasil pengolahan data yang dilakukan menggunakan Algoritma Random Forest, diperoleh beberapa simpulan antara lain Hasil yang diperoleh dengan melakukan percobaan yang berbeda-beda adalah nilai akurasi yang berbeda bahkan telah menunjukkan bahwa akurasi tertinggi diperoleh adalah 88,61%, Akurasi 88,61%, menunjukkan bahwa tingkat pendidikan di masyarakat, sangat mempengaruhi perilaku dan kepedulian terhadap kesehatan. Dari 790 data, terdapat 700 data yang memberikan trend positif pendidikan terhadap tingkat kesehatan masyarakat.

6. Daftar Pustaka

- [1] Agustiningrum, C. I., Gata, W., Nurfalah, R., & Radiah, U. (2020). *Komparasi Algoritma Random Forest, Random Forest Dan Svm Untuk Memprediksi Niat*. 20(2).
- [2] Imran, B., Subki, A., Yani, A., Alfian, M. R., Engineering, I., Engineering, S., & Mataram, U. T. (2019). *Data Mining Using Random Forest, Naïve Bayes, And Adaboost Models For Prediction And Classification Of Benign And*. 37–46.
- [3] Veronica Retno Sari, F. F. (2020). Perbandingan Prediksi Kualitas Kopi Arabika Dengan Menggunakan Algoritma Sgd, Random Forest Dan Naive Bayes. *EDUMATIC*, 4(2), 1–9.
- [4] Azhari, M. (2021). *Perbandingan Akurasi*,

- Recall , dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM, dan Naive Bayes. 5(April), 640–651.
- [5] Samosir, A., Hasibuan, M., Justino, W. E., & Hariyono, T. (2021). *Komparasi Algoritma Random Forest , Naïve Bayes dan K- Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung*. 214–222.
- [6] Rini, L. S. (2021). *Prediksi Penerima Beasiswa dengan Menggunakan Teknik Data Mining di Universitas Muhammadiyah Pringsewu*. 87–94.
- [7] Vimal, V. J. S., Mi, M. K., & Lee, Y. (2021). AI - based smart prediction of clinical disease using random forest classifier and Naive Bayes. *The Journal of Supercomputing*, 77(5), 5198–5219.
- [8] Ramdani. (2022). *Perbandingan Metode Random Forest Dan Random Forest Untuk Menentukan Prestasi Belajar Siswa Pada Jurusan RPL (Studi Kasus SMK Swasta Siti Banun Sigambal)*. 2, 80–88.
- [9] Nalatissifa, H., Gata, W., Diantika, S., & Nisa, K. (2021). *Perbandingan Kinerja Algoritma Klasifikasi Random Forest , Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja*. 5(4), 578–584.
- [10] Firmansyah, A. (2022). *Komparasi Akurasi Pada Naïve Bayes Dan Random Forest Dalam Klasifikasi Penyakit Liver Comparison of Accuracy in Naïve Bayes and Random Forests in Classification of Liver Disease*. 7(January), 81–89.
- [11] Alhabib, I., Faqih, A., & Dikananda, F. (2022). *Komparasi Metode Deep Learning , Naïve Bayes dan Random Forest untuk Prediksi Penyakit Jantung*. 6(2), 176–185.
- [12] Suandi, A. (2022). *Klasifikasi, Machine Learning, Naïve Bayes, Random Forest, PIP*. 10(2).
- [13] Irawan, D., Perkasa, E. B., Wahyuningsih, D., & Helmund, E. (2021). *Perbandingan Klasifikasi SMS Berbasis Support Vector Machine , Random Forest Classifier , Random Forest dan Bagging Classifier*. 10, 432–437.
- [14] Yahya, Nurhidayati and A. Suherman, “Prediksi Tingkat Kesehatan Masyarakat Berdasarkan Penggunaan Alat Kontrasepsi Menggunakan Algoritma Random Forest,” *Inform. dan Teknol.*, vol. 7, no. 1, 2024..
- [15] Putri, S. U. (n.d.). *Implementasi Data Mining Untuk Prediksi Penyakit Diabetes*. 2(1), 39–46.
- [16] Wijaya, Y. A., Bahtiar, A., Nining, R., Akuntansi, K., Tinggi, S., Informatika, M., & Cirebon, K. (2021). *Analisa Klasifikasi menggunakan Algoritma Decision Tree pada Data Log Firewall Jurnal Sistem Informasi dan Manajemen*. 9(3).
- [17] Hapsari, T. D., Malik, A., Simbolon, E. F., & Ariawan, I. (2022). *Classification of marine mammals based on nucleotides using machine learning*. 11(2), 62–70.
- [18] Tri widianti, J., Alfian, F. Y., & Prasojo, M. (2021). *Perbandingan Metode Data Mining Untuk Prediksi Prestasi Siswa Tingkat Pendidikan Menengah Kejuruan Pada Sekolah Menengah Kejuruan Negeri (SMKN 1) Gadingrejo Pringsewu Lampung*. 1(Smkn 1), 126–133.
- [19] Pinelis, M., & Ruppert, D. (2022). ScienceDirect. *The Journal of Finance and Data Science*, 8, 35–54.
- [20] Ath, S., Al, T., Darmawan, D., Fahmi, N., Hakim, A. , Qibtiyya, M. Al, & Syafei, N. S. (2022). *Jurnal Teknologi Terpadu Hybrid Machine Learning Model Untuk Memprediksi Penyakit Jantung Dengan Metode Logistic Regression Dan Random*. 8(1), 40–46.
- [21] Models, K., Nam, S., & Hur, J. (2018). *Probabilistic Forecasting Model of Solar Power Outputs Based on the Naïve Bayes Classifier and*.
- [22] Residencia, A. (2019). *Klasifikasi Kelayakan Peserta Pengajuan Kredit Rumah Dengan Algoritma Naïve Bayes Di Perumahan Azzura Residencia*. 9, 43–48