

IMPLEMENTASI ALGORITMA RANDOM FOREST UNTUK MENENTUKAN PENERIMA BANTUAN RASKIN

Ilham Kurniawan^{*1}, Duwi Cahya Putri Buani², Abdussomad³, Widya Apriliah⁴, Rizal Amegia Saputra⁵

^{1,3,4,5}Universitas Bina Sarana Informatika, Jakarta, ²Universitas Nusa Mandiri, Jakarta Pusat

Email: ¹ilham.imk@bsi.ac.id, ²duwi.dcp@nusamandiri.ac.id, ³abdussomad.bdu@bsi.ac.id,

⁴widya.wyr@bsi.ac.id, ⁵rizal.rga@bsi.ac.id

^{*}Penulis Korespondensi

(Naskah masuk: 18 Maret 2022, diterima untuk diterbitkan: 12 April 2023)

Abstrak

Kemiskinan adalah salah satu perhatian mendasar dari setiap pemerintah. Program Beras Keluarga Miskin (Raskin) merupakan salah satu program pemerintah. Skema raskin mempunyai tujuan meminimalisir beban rumah tangga tidak mampu sebagai bentuk bantuan untuk menaikkan ketahanan pangan melalui perlindungan sosial. Tujuan penelitian ini adalah menemukan akurasi tertinggi di antara algoritma klasifikasi prediktif yang diusulkan penerima bantuan raskin menggunakan *tools python machine learning* dan di implementasikan melalui suatu *website*. Klasifikasi adalah metode penambangan data yang menentukan kategori pada kelompok data untuk mendukung prediksi dan analisa yang semakin akurat. Beberapa algoritma klasifikasi pembelajaran mesin seperti, SVM, NB dan RF, digunakan pada penelitian ini demi menentukan penerima bantuan raskin. Eksperimen dilakukan menggunakan dataset Raskin Kelurahan Gunungparang, Kota Sukabumi yang bersumber dari Kelurahan Gunungparang. Kinerja algoritma klasifikasi dievaluasi dengan beragam metrik seperti *Precision*, *Accuracy*, *F-Measure*, dan *Recall*. Akurasi diukur melalui contoh yang dikelompokkan dengan benar atau salah. 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).

Kata kunci: *raskin, algoritma klasifikasi, random forest*

IMPLEMENTATION OF RANDOM FOREST ALGORITHM FOR DETERMINING RECIPIENTS OF RASKIN

Abstract

The problem of poverty is one of the fundamental concerns of every government. The Raskin program is one of the government's programs. The Raskin scheme has the aim of minimizing the burden on poor households in the form of assistance to improve food security by providing social protection. The purpose of this study is to find the highest accuracy among the predictive classification algorithms proposed by Raskin beneficiaries using python machine learning tools and implemented through a website. Classification is a data mining method that determines categories in data groups to support more accurate predictions and analysis. Therefore, three machine learning classification algorithms such as, support vector machine, naive bayes and random forest, were used in this experiment. to determine recipients of Raskin assistance. The experiment was carried out using the Raskin dataset, Gunungparang Village, Sukabumi City, which was sourced from Gunungparang Village. The performance of the classification algorithm is evaluated by various metrics such as Precision, Accuracy, F-Measure, and Recall. Accuracy is measured by correctly and incorrectly grouped samples. The results obtained show that the random forest classification algorithm has precision, recall, f-measure values with a value of 97%, an accuracy value of 97.26% and an ROC value of 0.970, better than other classification algorithms, namely the difference of 5.11% with the support vector classification algorithm. machine and 8.87% with naive bayes classification algorithm. Very good accuracy is used as a reference for algorithm performance if the number of False Negatives and False Positives is close. These results were proven accurately and systematically using Receiver Operating Characteristics (ROC).

Keywords: *raskin, classification algorithm, random forest*

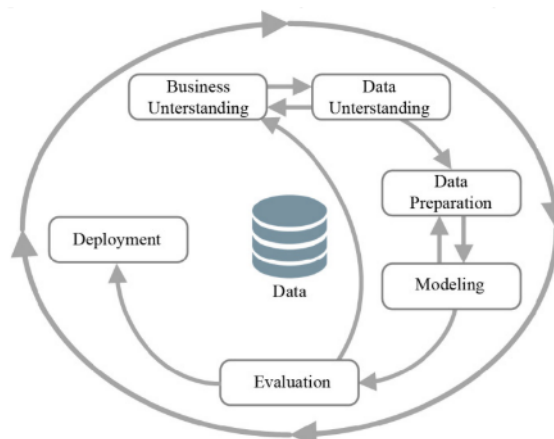
1. PENDAHULUAN

Kemiskinan menjadi salah satu persoalan yang dihadapi sebagian negara berkembang, termasuk negara Indonesia. Salah satu upaya untuk mengurangi kemiskinan, antara lain melalui program dukungan sosial, yaitu program beras keluarga miskin (Ermawati, 2019). Keberhasilan Rencana Raskin diukur dari tingkat pencapaian indikator yaitu: tepat sasaran, tepat jumlah, harga, waktu, kualitas dan tepat pengelolaan. Tugas kelurahan sangat berpengaruh terhadap kelancaran distribusi beras keluarga miskin, dan kelurahan harus mendistribusikannya kepada penerima manfaat beras keluarga miskin (Hidayah, 2019).

Pembelajaran mesin, diduga menjadi salah satu karakteristik terpenting dari kecerdasan artifisial, kecerdasan aritifisial mendukung peningkatan sistem komputer yang mampu mewarisi pengetahuan melalui pengalaman masa lalu tanpa diprogram untuk situasi tertentu. Pembelajaran mesin dianggap menjadi kepentingan mendesak dalam keadaan saat ini demi menghilangkan tenaga kerja manual melalui otomatisasi melalui sebuah *software* atau sistem informasi (Mujumdar dan Vaidehi, 2019). Data mining adalah tahap utama penemuan pengetahuan basis data, yang mencakup serangkaian metode untuk tujuan yang berbeda, seperti klasifikasi, pengelompokan, asosiasi, dan klasifikasi asosiasi (Alwidian, Hammo dan Obeid, 2018). Klasifikasi adalah salah satu fungsi utama dari *data mining*. Klasifikasi termasuk dalam *supervised learning* karena ada proses pembelajaran dari data masa lalu dalam proses klasifikasi. (Verawati dan Hasibuan, 2021). *Random forest* (RF) adalah evolusi *bagging*, yang bertujuan untuk mengurangi varians model statistik dan mensimulasikan variabilitas data dengan mengekstrak secara acak sampel bootstrap dari set pelatihan dan prediksi agregat untuk catatan baru. RF dapat meningkatkan prediksi banyak metode *supervised*, terutama pohon keputusan (Aria, Cuccurullo dan Gnasso, 2021). RF adalah salah satu algoritma klasifikasi *ensemble* berbasis pohon yang banyak digunakan. Banyak aspek integrasi pohon diperkenalkan untuk mengurangi korelasi antara pohon keputusan di hutan. *Bootstrap* digunakan dalam algoritma random forest untuk mengurangi pohon keputusan bias dan memutuskan pemisahan di setiap pohon keputusan (Mohana dkk., 2021). Tujuan dari penelitian ini diwakili oleh tiga hal objektif:

- Memperkenalkan serangkaian pendekatan yang ditujukan untuk memberikan interpretasi yang diperlukan untuk memahami model algoritma RF;
- Mengidentifikasi pendekatan terbaik untuk memberikan interpretasi model untuk model algoritma RF, melalui eksperimen komparatif.
- Mengimplemantasikan algoritma RF untuk menentukan penerima bantuan raskin.

2. METODE PENELITIAN



Gambar 1. CRISP-DM (Huber dkk., 2019)

Untuk mengkonfirmasi pandangan para ahli, kami juga memeriksa bahwa CRISP-DM masih merupakan metode yang sangat umum dalam aplikasi data mining. Sebagai contoh, hanya berfokus pada empat tahun terakhir, kita dapat menemukan sejumlah besar studi tradisional yang menerapkan atau sedikit menyesuaikan metode CRISP-DM ke berbagai bidang seperti, kesehatan (Contreras dkk., 2018), pemrosesan sinyal (Daderman dan Rosander, 2018), engineering (Rogalewicz dan Sika, 2016) dan (Huber dkk., 2019), pendidikan (Almahadeen, Akkaya dan Sari, 2017) dan (Castro R., Espitia P. dan Montilla, 2018), sensor dan aplikasi yang dapat digunakan (Nagashima dan Kato, 2019), olahraga (Bunker dan Thabtah, 2019) dan hukum (Barros dkk., 2018). Metode penelitian yang dilakukan pada penelitian ini yang dapat dilihat pada gambar 1 adalah sebagai berikut :

1. Bussiness Understanding
2. Data Understanding
3. Data Preparation
4. Modelling
5. Evaluation
6. Deployment

Yang akan dijelaskan pada point-point berikut :

2.1. Bussiness Understanding

Langkah awal ini berpusat pada penafsiran tujuan dan persyaratan sisi bisnis, yang kemudian diterjemahkan ke dalam pengetahuan untuk mendefinisikan definisi masalah utama yang dapat diatasi melalui *data mining*. (Andriawan dkk., 2020).

2.2. Data Understanding

Tahap ini berupa peneliti melakukan proses mengidentifikasi data, memahami kualitas data, mendapatkan wawasan awal dari data, dan memperoleh beberapa hipotesis untuk menemukan informasi yang tersembunyi di dalam data. Pada tahap ini, visualisasi data dilakukan untuk memahami data dan membersihkannya dari data yang hilang atau menghapus fitur yang bermasalah untuk menghasilkan

model pembelajaran mesin yang lebih baik dan lebih umum. (Pinto dkk., 2020).

2.3. Data Preparation

Pada tahap ini, semua kegiatan adalah membuat dataset akhir, yang akan dimasukkan ke dalam model yang dibuat. Persiapan data melibatkan semua aktivitas membangun kumpulan data untuk digunakan dalam sebuah model (Gonçalves dkk., 2020).

Tabel 1. Konversi dataset

| Jenis_Lantai_Rumahan | Jenis_Dinding_Rumahan | Sumber_Penerangan_Utama | Pekerjaan | Keperluan_Aset | Bahan_Bakar_Memasa | Hasil |
|----------------------|-----------------------|-------------------------|-----------|----------------|--------------------|-------|
| 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 1 | 0 | 0 | 4 | 1 | 0 | 0 |
| 2 | 1 | 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 0 | 4 | 1 | 0 | 0 |
| 1 | 0 | 0 | 4 | 1 | 0 | 0 |

Tabel 2 merupakan sampel dataset yang sudah di konversikan dari format non numerik yang disajikan pada Tabel 1 menjadi format numerik. Untuk keperluan pemrosesan data, seperti pada machine learning, data harus direpresentasikan dalam format numerik agar dapat diproses menggunakan rumus matematika.

2.4. Modelling

Langkah pertama dalam proses pemodelan adalah memilih calon model yang akan digunakan dalam eksperimen. Fase ini akan melibatkan peninjauan literatur masa lalu dan mengidentifikasi model prediksi yang umum digunakan sebelumnya yang sukses (Bunker dan Thabtah, 2019).

2.5. Evaluation

Pada tahap ini model dievaluasi dan kami juga menganalisis beberapa langkah yang dilakukan sebelumnya untuk memastikan bahwa model yang dipilih memiliki kualitas terbaik yang dapat mencapai tujuan dari masalah bisnis yang ada (van der Voort dkk., 2021).

2.6. Deployment

Penerapan model (*Deployment*) adalah tahap terakhir dan paling menantang dari siklus hidup pembelajaran mesin. Hal ini dikarenakan model dari hasil pengujian akan diimplementasikan dalam aplikasi nyata, menggunakan data nyata di lapangan.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan pembelajaran mesin python menggunakan laptop dengan prosesor Intel Corei5 RAM 4 Giga Byte hardisk 500 Giga Byte.

3.1. Business Understanding

Menurut data BPS 2009, nyaris seperempat ataupun 24% pada total masyarakat Indonesia dinyatakan tidak mampu. (Yesnita, 2014). Program Raskin dianggap tercapai ketika metrik (6T) terpenuhi,

yaitu tepat jumlah, tepat sasaran, tepat waktu, tepat harga, tepat kualitas, dan tepat pengelolaan. (Astuti, 2014). Diseminasi Program Raskin kepada rakyat selama ini minim, sehingga misi dan sasaran program raskin nyaris serupa diseluruh daerah, merupakan kurang tepat sasaran. (Widowati, 2015). Oleh karena itu, prediksi yang baik dengan model pembelajaran mesin dapat membantu untuk mengambil langkah-langkah preventif dalam menangani masalah-masalah yang terjadi.

3.2. Data Understanding

Data yang kami gunakan yaitu data yang diperoleh dari kelurahan Gunungparang, dimana pada data tersebut terdapat data warga yang berhak dan tidak berhak mendapat Pemberian beras keluarga miskin, warga yang berhak dan tidak berhak menerima bantuan sudah ditentukan sebelumnya oleh perangkat desa Gunungparang dengan berdasarkan variable-variabel yang sudah ditetapkan oleh kementrian sosial. Data disajikan sebagai daftar model dan variabel yang masing-masing merupakan data masyarakat Kelurahan Gunungparang sejumlah 418 masyarakat, terdiri dari 259 masyarakat yang mempunyai hak mendapatkan bantuan dan 159 masyarakat yang tidak mempunyai hak mendapatkan bantuan. Sebelum langkah pemodelan, dataset perlu dibersihkan dan juga dimodifikasi.

Tabel 2. Sampel Dataset penerima bantuan raskin

| Jenis_Lantai_Rumahan | Jenis_Dinding_Rumahan | Sumber_Penerangan_Utama | Pekerjaan | Keperluan_Aset | Bahan_Bakar_Memasa | Hasil |
|----------------------|-----------------------|-------------------------|---------------------|----------------|--------------------|-------|
| Semen/Papan | Bambu | Listrik | Pekerja Lepas | Tidak | Arang/Kayu | Laya |
| Tanah | Bambu | Listrik | Tidak/Belum Bekerja | Tidak | Arang/Kayu | Laya |
| Ubin/Keramik/Marmar | Tembok | Listrik | Pedagang | Memiliki | Listrik/Gas | Tidak |
| Semen/Papan | Bambu | Listrik | Tidak/Belum Bekerja | Tidak | Arang/Kayu | Laya |
| Tanah | Bambu | Listrik | Tidak/Belum Bekerja | Tidak | Arang/Kayu | Laya |

Tabel 2 merupakan sampel data dari dataset yang digunakan untuk memprediksi penerima bantuan raskin. Data tersebut merupakan data dengan format non numerik.

3.3. Data Preparation

Penyusunan data dalam hal ini akan fokus pada pemilihan fitur yang digunakan dalam model, kami juga akan melakukan beberapa rekayasa fitur. Penyusunan data akan difokuskan pada pemilihan fitur yang digunakan dalam model dan kami juga melakukan

beberapa rekayasa fitur. Fitur akhir yang digunakan pada model terdiri dari 7 fitur. Dataset yang peneliti gunakan pada penelitian ini yaitu data non numerik yang akan dikodekan menggunakan teknik *one hot encoding*.

3.4. Modelling

Model yang kami gunakan adalah algoritma klasifikasi yaitu, *Random Forest* (RF), *Support Vector Machine* (SVM) dan *Naive Bayes* (NB). Peneliti memutuskan untuk menggunakan algoritma berbasis pohon dalam memecahkan masalah prediksi dalam penentuan penerima bantuan raskin karena terbukti bekerja dengan baik pada penelitian sebelumnya (Mujumdar dan Vaidehi, 2019) dan (Apriliyah dkk., 2021) (Makariou, Barrieu dan Chen, 2021), beberapa model yang akan digunakan antara lain:

- 1) RF, Metode *ensemble* yang kami gunakan disebut *random forest*. Ini dikembangkan oleh Breiman (2001) dan digunakan untuk memecahkan masalah prediksi. RF adalah model pembelajaran mesin yang menggabungkan beberapa pohon keputusan menjadi satu model saja. Hutan Acak tergantung pada nilai vektor acak yang memiliki distribusi yang sama di setiap pohon di setiap pohon keputusan dan memiliki kedalaman maksimum (Makariou, Barrieu dan Chen, 2021). RF adalah *ensemble learner*, sebuah metode yang menghasilkan beberapa *classifier* dan menggabungkan hasilnya menjadi satu model (Jain dan Saha, 2021).
- 2) SVM
SVM adalah algoritma klasifikasi berbasis diskriminatif yang bermaksud untuk mencari batasan pemisahan terbaik yang disebut *hyperplane* yang akan memisahkan kelas. Contoh yang paling dekat pada *hyperplane* disebut *support vector*, dan selisihnya dinyatakan seperti jumlah berat melalui subset contoh yang membatasi kompleksitas masalah (Kurniawan, 2020). SVM bekerja dengan menginterpolasi *hyperplane* antar kelas dan membimbingnya dengan cara yang menjaga jarak maksimal dari titik data. (Shankar dkk., 2020).
- 3) NB
NB adalah algoritma klasifikasi yang mengaplikasikan prinsip *Bayesian* pada algoritma klasifikasi (Moreira dan Namen, 2018). Teorema keputusan *Bayesian* merupakan cara statistik dasar pemahaman pola. NBC didasarkan pada dugaan sederhana bahwa, dengan nilai keluaran, nilai atribut secara tentatif independen satu sama lain. Dengan istilah lain, diberikan jumlah *output*, probabilitas pengamatan kolektif merupakan hasil dari probabilitas pribadi (Tigga dan Garg, 2020).

Pengujian dilakukan dengan memakai teknik *10-fold cross validation* terhadap algoritma klasifikasi RF, SVM dan NB, sehingga menghasilkan *confussion matrix*. *Confussion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu

metode klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya.

Tabel 3. *Confussion Matrix*

| Algoritma | TP | FN | FP | TN |
|-----------|----|----|----|----|
| RF | 74 | 1 | 3 | 48 |
| SVM | 65 | 1 | 12 | 48 |
| NB | 64 | 13 | 0 | 49 |

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

3.5. Evaluation

Evaluasi didasarkan pada nilai akurasi, *recall*, presisi, dan *f-measure* yang dihasilkan oleh masing-masing model. Penulis membuat model untuk memprediksi variabel target yaitu hasil yang merupakan nilai biner (0 berarti layak menerima bantuan raskin, dan 1 jika tidak layak), beberapa algoritma yang digunakan penulis antara lain RF, SVM dan NB. Penulis menggunakan *K-fold 10-Cross Validation*, suatu teknik evaluasi model yang cukup populer dan banyak digunakan. Penulis memilih 10 untuk nilai K. Artinya, untuk setiap iterasi, algoritma K-fold mengevaluasi model menggunakan 10% data uji dan 90% data latih. Setelah didapatkan model terbaik, selanjutnya dilakukan *fine-tuning* model *hyperparameter* untuk mendapatkan nilai terbaik.

Berdasarkan nilai TN, FP, FN, dan TP pada Tabel 3 dapat diperoleh nilai *accuracy*, *precision*, *recall* dan *f-measure*. Nilai akurasi menggambarkan seberapa akurat sistem mengklasifikasikan data dengan benar.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (1)$$

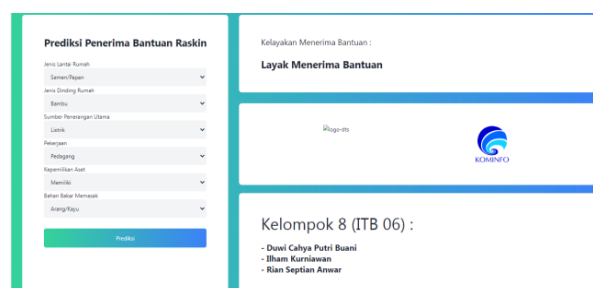
$$\text{Precision} = \frac{TP}{FP+TP} \times 100\% \quad (2)$$

$$\text{Recall} = \frac{TP}{FN+TP} \times 100\% \quad (3)$$

$$\text{F-Measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (4)$$

3.6. Deployment

Peneliti mengimplementasikan penerimaan bantuan raskin melalui algoritma klasifikasi dengan menerapkan dalam sebuah sistem informasi yang dapat diakses melalui [link https://tugas-akhir-kelompok08.herokuapp.com/](https://tugas-akhir-kelompok08.herokuapp.com/).



Gambar 2. *Deployment model*

3.7. Pengukuran Akurasi

Penelitian ini memanfaatkan algoritma klasifikasi NB, SVM dan RF. Pengujian dilakukan dengan memakai teknik *10-fold cross validation*. Pengukuran akurasi, *f-measure*, *recall*, *precision*, dan ROC (*Receiver Operating Curve*) dimanfaatkan bagi klasifikasi dalam penelitian ini.

Tabel 4. Perbandingan kinerja algoritma klasifikasi pada berbagai metrik

| Algoritma | Precision | Recall | F-Measure | Accuracy % | ROC |
|-----------|-----------|--------|-----------|------------|-------|
| RF | 0,970 | 0,970 | 0,970 | 97,26 | 0,970 |
| SVM | 0,910 | 0,900 | 0,900 | 92,15 | 0,911 |
| NBC | 0,920 | 0,900 | 0,900 | 88,39 | 0,915 |

Tabel 4 menunjukkan angka hasil yang beragam untuk semua algoritma klasifikasi yang dihitung melalui beragam metrik. Dapat dilihat dari Tabel 4 algoritma klasifikasi RF memiliki nilai *precision*, *recall*, *f-measure* dengan nilai 97%, nilai *accuracy* sebesar 97,26% dan nilai ROC 0,970, mengungguli algoritma klasifikasi lainnya. *F-measure* menggambarkan perbandingan rata-rata *precision* dan *recall*. Jika dataset memiliki *false negative* dan *false positive* yang sangat dekat (*symmetric*), kami menggunakan akurasi sebagai acuan untuk kinerja algoritma. Namun, jika angkanya tidak dekat, maka harus menggunakan *f-measure* sebagai acuan. Berdasarkan Tabel 3 nilai FN dan FP algoritma klasifikasi RF memiliki jumlah yang sangat dekat sehingga nilai *accuracy* algoritma RF bisa dijadikan acuan, algoritma RF dinilai baik dalam mengklasifikasi dan melakukan prediksi penerimaan bantuan Raskin. Nilai ROC merepresentasikan algoritma klasifikasi yang dibangun dalam memprediksi objek penerima bantuan Raskin, jika nilai ROC semakin dekat dengan angka 1 maka algoritma klasifikasi yang dibangun semakin baik pula, algoritma klasifikasi RF bisa memprediksi objek penerima bantuan Raskin lebih akurat dibandingkan algoritma klasifikasi yang diusulkan lainnya.

4. KESIMPULAN DAN SARAN

Bersumber pada penelitian yang sudah dilakukan, maka dapat ditarik kesimpulan penerapan metode *Random Forest* dapat mengklasifikasikan data penerima bantuan raskin dengan benar dengan performa yang sangat baik yaitu *accuracy* sebesar 97,26%, dan ROC sebesar 0,970 dibandingkan dengan algoritma SVM dan NB yang memiliki perbedaan perbedaan sebesar 5,11% dengan algoritma klasifikasi SVM dan 8,87% dengan algoritma klasifikasi NB, dapat dibangunnya sebuah *website* sistem informasi yang dapat menentukan penerima bantuan raskin sesuai dengan kriteria yang sudah ada dengan algoritma klasifikasi *Random Forest*. Saran yang bisa peneliti sampaikan yaitu lebih banyak data pelatihan ditambahkan sehingga model dapat

memberikan kinerja yang lebih baik, dan menggunakan metode optimasi seperti *Particle Swarm Optimization* untuk melihat perbedaannya tanpa menggunakan metode optimasi.

DAFTAR PUSTAKA

- ALMAHADEEN, L., AKKAYA, M. dan SARI, A., 2017. Mining Student Data Using CRISP- DM Model. *International Journal of Computer Science and Information Security*.
- ALWIDIAN, J., HAMMO, B.H. dan OBEID, N., 2018. WCBA: Weighted classification based on association rules algorithm for breast cancer disease. *Applied Soft Computing Journal*, [daring] 62, hal.536–549. <https://doi.org/10.1016/j.asoc.2017.11.013>.
- ANDRIAWAN, Z.A., PURNAMA, S.R., DARMAWAN, A.S., RICKO, WIBOWO, A., SUGIHARTO, A. dan WIJAYANTO, F., 2020. Prediction of Hotel Booking Cancellation using CRISP-DM. *ICICoS 2020 - Proceeding: 4th International Conference on Informatics and Computational Sciences*, hal.0–5. <https://doi.org/10.1109/ICICoS51170.2020.9299011>.
- APRILIAH, W., KURNIAWAN, I., BAYDHOWI, M. dan HARYATI, T., 2021. Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. *Sistemasi*, 10(1), hal.163. <https://doi.org/10.32520/stmsi.v10i1.1129>.
- ARIA, M., CUCCURULLO, C. DAN GNASSO, A., 2021. A comparison among interpretative proposals for Random Forests. *Machine Learning with Applications*, [daring] 6(June), hal.100094. <https://doi.org/10.1016/j.mlwa.2021.100094>.
- BARROS, R., PERES, A., LORENZI, F., KRUG WIVES, L. dan HUBERT DA SILVA JACCOTTET, E., 2018. Case law analysis with machine learning in Brazilian court. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10868 LNAI, hal.857–868. https://doi.org/10.1007/978-3-319-92058-0_82.
- BUNKER, R.P. dan THABTAH, F., 2019. A machine learning framework for sport result prediction. *Applied Computing and Informatics*, [daring] 15(1), hal.27–33. <https://doi.org/10.1016/j.aci.2017.09.005>.
- CASTRO R., L.F., ESPITIA P., E. dan MONTILLA, A.F., 2018. Applying CRISP-DM in a KDD process for the analysis of student attrition. *Communications in Computer and Information Science*, 885, hal.386–401.

- https://doi.org/10.1007/978-3-319-98998-3_30.
- CONTRERAS, Y., VERA, M., HUÉRFANO, Y., VALBUENA, O., SALAZAR, W., VERA, M.I., BORRERO, M., BARRERA, D., HERNÁNDEZ, C., MOLINA, Á.V., MARTÍNEZ, L.J., SÁENZ, F., VIVAS, M., SALAZAR, J. dan GELVEZ, E., 2018. Digital processing of medical images: Application in synthetic cardiac datasets using the CRISP_DM methodology. *Revista Latinoamericana de Hipertension*, 13(4), hal.310–315.
- DADERMAN, A. dan ROSANDER, S., 2018. Evaluating Frameworks for Implementing Machine Learning in Signal Processing. *Examensarbete Inom Teknik*, hal.1–36.
- ERMAWATI, E., 2019. Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Prediksi Penerima Bantuan Pangan Non Tunai. *Sistmasi*, 8(3), hal.513. <https://doi.org/10.32520/stmsi.v8i3.576>.
- GONÇALVES, C., FERREIRA, D., NETO, C., ABELHA, A. dan MACHADO, J., 2020. Prediction of mental illness associated with unemployment using data mining. *Procedia Computer Science*, [daring] 177, hal.556–561. <https://doi.org/10.1016/j.procs.2020.10.078>.
- HIDAYAH, N., 2019. Sistem Klasifikasi Penerima Beras Miskin Menggunakan Algoritma Decision Tree C4.5. *Program Studi Teknik Informatika, Fakultas Teknologi Informasi & Elektro, Universitas Teknologi Yogyakarta*, 5.
- HUBER, S., WIEMER, H., SCHNEIDER, D. dan IHLENFELDT, S., 2019. DMME: Data mining methodology for engineering applications - A holistic extension to the CRISP-DM model. *Procedia CIRP*, [daring] 79, hal.403–408. <https://doi.org/10.1016/j.procir.2019.02.106>.
- JAIN, S. dan SAHA, A., 2021. Improving performance with hybrid feature selection and ensemble machine learning techniques for code smell detection. *Science of Computer Programming*, [daring] 212, hal.102713. <https://doi.org/10.1016/j.scico.2021.102713>.
- KURNIAWAN, I., 2020. Prediksi Gejala Autism Spectrum Disorders pada Remaja Menggunakan Optimasi Particle Swarm Optimization dan Algoritma Support Vector Machine. *INFORMATICS FOR EDUCATORS AND PROFESSIONALS*, 4(2), hal.113–122.
- MAKARIOU, D., BARRIEU, P. dan CHEN, Y., 2021. A random forest based approach for predicting spreads in the primary catastrophe bond market. *Insurance: Mathematics and Economics*, [daring] 101, hal.140–162. <https://doi.org/10.1016/j.insmatheco.2021.07.003>.
- MOHANA, R.M., REDDY, C.K.K., ANISHA, P.R. dan MURTHY, B.V.R., 2021. Random forest algorithms for the classification of tree-based ensemble. *Materials Today: Proceedings*, [daring] (xxxx). <https://doi.org/10.1016/j.matpr.2021.01.788>.
- MOREIRA, L.B. dan NAMEN, A.A., 2018. A hybrid data mining model for diagnosis of patients with clinical suspicion of dementia. *Computer Methods and Programs in Biomedicine*, [daring] 165, hal.139–149. <https://doi.org/10.1016/j.cmpb.2018.08.016>.
- MUJUMDAR, A. dan VAIDEHI, V., 2019. Diabetes Prediction using Machine Learning Algorithms. *Procedia Computer Science*, [daring] 165, hal.292–299. <https://doi.org/10.1016/j.procs.2020.01.047>.
- NAGASHIMA, H. dan KATO, Y., 2019. APREP-DM: A Framework for Automating the Pre-Processing of a Sensor Data Analysis based on CRISP-DM. *2019 IEEE International Conference on Pervasive Computing and Communications Workshops, PerCom Workshops 2019*, hal.555–560. <https://doi.org/10.1109/PERCOMW.2019.8730785>.
- PINTO, A., FERREIRA, D., NETO, C., ABELHA, A. dan MACHADO, J., 2020. Data mining to predict early stage chronic kidney disease. *Procedia Computer Science*, 177(2018), hal.562–567. <https://doi.org/10.1016/j.procs.2020.10.079>.
- ROGALEWICZ, M. dan SIKAR, R., 2016. Methodologies of knowledge discovery from data and data mining methods in mechanical engineering. *Management and Production Engineering Review*, 7(4), hal.97–108. <https://doi.org/10.1515/mper-2016-0040>.
- SHANKAR, K., LAKSHMANAPRABU, S.K., GUPTA, D., MASELENO, A. dan DE ALBUQUERQUE, V.H.C., 2020. Optimal feature-based multi-kernel SVM approach for thyroid disease classification. *Journal of Supercomputing*, [daring] 76(2), hal.1128–1143. <https://doi.org/10.1007/s11227-018-2469-4>.
- TIGGA, N.P. dan GARG, S., 2020. Prediction of Type 2 Diabetes using Machine Learning Classification Methods. *Procedia Computer Science*, [daring] 167(2019), hal.706–716. <https://doi.org/10.1016/j.procs.2020.03.336>.
- VERAWATI, Y. dan HASIBUAN, M.S., 2021. Perbandingan Data Set IRIS Dengan Aplikasi Rapid Miner dan Orange menggunakan Algoritma Klasifikasi. *Institut Informatika dan Bisnis Darmajaya*, hal.158–163.
- VAN DER VOORT, H., VAN BULDEREN, S., CUNNINGHAM, S. dan JANSSEN, M., 2021.

Data science as knowledge creation a framework for synergies between data analysts and domain professionals. *Technological Forecasting and Social Change*, [daring] 173(January), hal.121160. <https://doi.org/10.1016/j.techfore.2021.121160>

Halaman ini sengaja dikosongkan