

**IMPLEMENTASI ALGORITMA RANDOM FOREST DALAM
KLASIFIKASI KELAYAKAN MASYARAKAT UNTUK
MENDAPATKAN BANTUAN SOSIAL
(STUDI KASUS : DESA ALANG BON-BON, KABUPATEN
ASAHAN)**

SKRIPSI

OLEH:

**PRANS HARDIANSYAH SIMANJUNTAK
188160059**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2024**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 16/1/25

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

**IMPLEMENTASI ALGORITMA RANDOM FOREST DALAM
KLASIFIKASI KELAYAKAN MASYARAKAT UNTUK
MENDAPATKAN BANTUAN SOSIAL
(STUDI KASUS : DESA ALANG BON-BON, KABUPATEN
ASAHAN)**

SKRIPSI

Diajukan sebagai Salah Satu Syarat untuk Memperoleh
Gelar Sarjana di Fakultas Teknik
Universitas Medan Area



OLEH:

**PRANS HARDIANSYAH SIMANJUNTAK
188160059**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
MEDAN
2024**

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 16/1/25

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

HALAMAN PENGESAHAN

Judul Skripsi : Implementasi Algoritma *Random Forest* Dalam
Klasifikasi Kelayakan Masyarakat Untuk Mendapatkan
Bantuan Sosial (Studi Kasus: Desa Alang Bon – Bon,
Kabupaten Asahan)

Nama : Prans Hardiansyah Simanjuntak

NPM : 188160059

Fakultas : Teknik

Disetujui Oleh
Komisi Pembimbing


Andre Hasudungan Lubis, S. Ti, M. Sc
Pembimbing


Dr. Eng. Supriatno, ST, MT
Dekan Fakultas Teknik


Rizki Muliono, S. Kom, M. Kom
Ka. Prodi

Tanggal Lulus : 01 September 2024

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian – bagian tertentu dalam penulisan skripsi ini yang saya kutip dari karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksi – sanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan, 01 September 2024



Prans Hardiansyah Simanjuntak
188160059

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Prans Hardiansyah Simanjuntak
NPM : 188160059
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, memyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Non eksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul : **Implementasi Algoritma *Random Forest* Dalam Klasifikasi Kalayakan Masyarakat Untuk Mendapatkan Bantuan Sosial (Studi Kasus: Desa Alang Bon – Bon, Kabupaten Asahan)**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Dibuat di : Medan
Pada Tanggal : 01 September 2024
Yang Menyatakan



(Prans Hardiansyah Simanjuntak)



RIWAYAT HIDUP

Penulis lahir pada tanggal 19 Juni 2000 di Desa Alang Bon Bon Kecamatan Aek Kuasan Kabupaten Asahan. Penulis merupakan Anak ke 5 dari 6 bersaudara dari Ayah Johan Simanjuntak dan Ibu Uji Usni Br Panjaitan.

Pendidikan yang pernah ditempuh penulis sampai saat ini adalah : sekolah Dasar (SDN) 015930 Alang Bon Bon Kecamatan Aek Kuasan Kabupaten Asahan lulus tahun 2012, Sekolah Menengah Pertama (SMPN2) Alang Bon Bon Kecamatan Aek Kuasan Kabupaten Asahan lulus pada tahun 2015, Sekolah Menengah Kujuruan (SMK) Swasta Pulau Rakyat lulus pada tahun 2018 dan pada tahun 2018 terdaptar sebagai Mahasiswa Teknik Informatika Fakultas Teknik Universitas Medan Area.

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “**Implementasi Algoritma *Random Forest* Dalam Klasifikasi Kelayakan Masyarakat Untuk Mendapatkan Bantuan Sosial (Studi Kasus: Desa Alang Bon – Bon, Kabupaten Asahan)**” ini dengan baik. Skripsi ini disusun untuk memenuhi salah satu syarat menyelesaikan pendidikan Program Strata-1 pada Fakultas Teknik Program Studi Teknik Informatika di Universitas Medan Area.

Dalam proses menyelesaikan skripsi ini, penulis menyadari bahwa tulisan ini masih jauh dari kata sempurna dan juga terdapat banyak kekurangan. Penulis juga mengharapkan kritik dan saran yang bersifat membangun dari para pembaca. Kemudian penulis ingin mengucapkan terima kasih kepada:

1. Tuhan Yang Maha Esa, sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
2. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc, selaku Rektor Universitas Medan Area.
3. Bapak Dr. Eng. Supriatno, ST, MT, selaku Dekan Fakultas Teknik Universitas Medan Area.
4. Bapak Rizki Muliono, S.Kom, M.Kom, selaku Kepala Program Studi Teknik Informatika.
5. Bapak Andre Hasudungan Lubis, S. Ti, M. Sc, selaku Dosen Pembimbing yang telah mengarahkan, membimbing, dan memberikan motivasi kepada penulis dalam proses menyelesaikan skripsi ini.
6. Seluruh Dosen Teknik Informatika Universitas Medan Area yang selama ini telah membekali penulis dengan ilmu yang sangat bermanfaat.
7. Seluruh Pegawai Universitas Medan Area yang telah membantu dalam

proses administrasi.

8. Teman – teman Teknik Informatika 2018, terima kasih atas persaudaraannya selama ini.

Penulis menyadari dalam penyusunan skripsi ini, masih terdapat banyak kekurangan dan jauh dari kesempurnaan, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya.

Medan, 01 September 2024

Prans Hardiansyah Simanjuntak
188160059



ABSTRAK

Kelayakan masyarakat untuk mendapatkan bantuan sosial merupakan salah satu masalah yang dialami oleh beberapa negara berkembang, termasuk Indonesia. Dikarenakan banyaknya penduduk, lapangan pekerjaan terbatas, yang mengakibatkan banyak pengangguran. Desa Alang Bon-Bon merupakan salah satu wilayah yang berada di Kecamatan Aek Kuasan yang memiliki jumlah RW (Rukun Warga) sebanyak 11 RW dimana di tiap RW tersebut memiliki beberapa RT (Rukun Tetangga) dengan jumlah penduduk kurang lebih 2719 jiwa. Dari data masyarakat kurang mampu tersebut, penerima bantuan sosial sebanyak 207 KK. Dengan adanya masalah tersebut, mengakibatkan adanya ketidak sinkronan penerima bantuan sosial. Masyarakat yang berhak menerima bisa jadi tidak menerima, sedangkan warga yang bukan merupakan masyarakat kurang mampu malah justru menerima bantuan sosial. Klasifikasi adalah proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Salah satu metode klasifikasi yang bisa digunakan adalah *Random Forest* yang merupakan dapat meningkatkan hasil akurasi, karena dalam membangkitkan simpul anak untuk setiap *node* dilakukan secara acak. Hasil terbaik yaitu pada prosedur ke-3 dengan menggunakan *data training* 80% dan *data testing* 20% mendapatkan akurasi sebesar 97%, *precision* 97%, *recall* 97%, dan *f1-score* 97% sehingga mendapatkan performa yang tinggi serta prediksinya yang akurat.

Kata Kunci : *Random Forest*, Klasifikasi, Bantuan Sosial, Desa Alang Bon-Bon

ABSTRACT

People's eligibility for social assistance is one of the problems experienced by several developing countries, including Indonesia. Due to the large population, employment opportunities are limited, which results in a lot of unemployment. Alang Bon-Bon Village is one of the areas in Aek Kuasan Sub-district which has a total of 11 RW (Rukun Warga) where each RW has several RT (Rukun Tetangga) with a population of approximately 2719 people. From the data on the underprivileged community, the recipients of social assistance are 207 families. With this problem, there is a synchronization of social assistance recipients. People who are entitled to receive it may not receive it, while residents who are not poor people actually receive social assistance. Classification is the process of finding models or functions that describe and distinguish classes of data or concepts that aim to be used to predict the class of objects whose class labels are unknown. One of the classification methods that can be used is Random Forest which can improve accuracy results, because in generating child nodes for each node is done randomly. The best results are in the 3rd procedure using 80% training data and 20% testing data to get 97% accuracy, 97% precision, 97% recall, and 97% f1-score so as to get high performance and accurate predictions.

Keywords: Random Forest, Classification, Social Assistance, Desa Alang Bon-Bon

DAFTAR ISI

	Halaman
RIWAYAT HIDUP	v
KATA PENGANTAR	vi
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	5
2.1 Kelayakan Bantuan Sosial	5
2.2 Program Kelayakan Masyarakat (PKM)	5
2.3 Klasifikasi	6
2.4 <i>Random Forest</i>	6
2.5 <i>Python</i>	8
2.6 <i>Confusion Matrix</i>	9
2.7 Penelitian Terdahulu	11
BAB III METODOLOGI PENELITIAN	17
3.1 Prosedur Penelitian	17
3.2 Data dan Sumber Data	18
3.3 Teknik Analisis Data	18
3.4 Implementasi <i>Random Forest</i>	19
BAB IV HASIL DAN PEMBAHASAN	29
4.1 Pengumpulan Data	29
4.2 <i>Data Preprocessing</i>	30
4.3 Normalisasi Data	32
4.4 Implementasi	33
4.5 Hasil	35

4.6	Evaluasi	37
4.7	Pembahasan	40
BAB V KESIMPULAN DAN SARAN		41
5.1	Kesimpulan	41
5.2	Saran	41
DAFTAR PUSTAKA		42



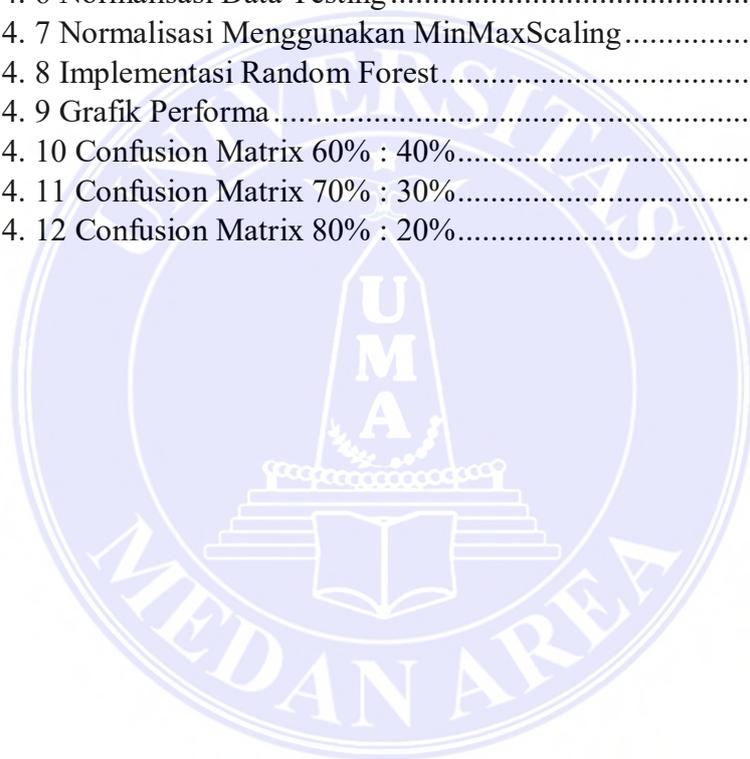
DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	9
Tabel 2. 2 Penelitian Terdahulu.....	11
Tabel 3. 1 Proporsi Data.....	18
Tabel 4. 1 Keterangan Hasil Encoding	31
Tabel 4. 2 Hasil Klasifikasi Data Training 60% dan Testing 40%.....	34
Tabel 4. 3 Hasil Klasifikasi Data Training 70% dan Testing 30%.....	34
Tabel 4. 4 Hasil Klasifikasi Data Training 80% dan Testing 20%.....	35
Tabel 4. 5 Performa	35



DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Metode Random Forest	7
Gambar 3. 1 Prosedur Penelitian	17
Gambar 3. 2 Implementasi Random Forest.....	19
Gambar 3. 3 Data Latih.....	20
Gambar 3. 7 Pohon Keputusan.....	28
Gambar 4. 1 Syntax Membaca Dataset.....	29
Gambar 4. 2 Pengumpulan Data.....	29
Gambar 4. 3 Label Encoding.....	30
Gambar 4. 4 Syntax Label Encoding	31
Gambar 4. 5 Normalisasi Data Training	32
Gambar 4. 6 Normalisasi Data Testing.....	32
Gambar 4. 7 Normalisasi Menggunakan MinMaxScaling.....	33
Gambar 4. 8 Implementasi Random Forest.....	33
Gambar 4. 9 Grafik Performa.....	36
Gambar 4. 10 Confusion Matrix 60% : 40%.....	37
Gambar 4. 11 Confusion Matrix 70% : 30%.....	38
Gambar 4. 12 Confusion Matrix 80% : 20%.....	39



BAB I

PENDAHULUAN

1.1 Latar Belakang

Kelayakan masyarakat untuk mendapatkan bantuan sosial merupakan salah satu masalah yang dialami oleh beberapa negara berkembang, termasuk Indonesia. Dikarenakan banyaknya penduduk, lapangan pekerjaan terbatas, yang mengakibatkan banyak pengangguran (Pratama, 2022). Penggolongan kemiskinan didasarkan pada suatu standar tertentu yaitu dengan membandingkan tingkat pendapatan orang atau keluarga dengan tingkat pendapatan yang diperlukan untuk memenuhi kebutuhan pokok minimum. Dalam mengatasi permasalahan tersebut, pemerintah tentu tidak tinggal diam. Berbagai upaya telah dilakukan untuk mengatasi kemiskinan, salah satunya dengan adanya program bantuan sosial yang berada dibawah koordinasi Tim Koordinasi Penanggulangan Kemiskinan (TKPK), baik dipusat maupun didaerah. PKM adalah program perlindungan sosial yang memberikan bantuan tunai kepada Rumah Tangga Sangat Miskin (RTSM). Program ini dalam jangka pendek bertujuan mengurangi beban RTSM dan dalam jangka panjang diharapkan dapat memutus mata rantai kemiskinan antar generasi, sehingga generasi berikutnya dapat keluar dari perangkap kemiskinan (Hermansyah, 2022).

Desa Alang Bon-Bon merupakan salah satu wilayah yang berada di Kecamatan Aek Kuasan yang memiliki jumlah RW (Rukun Warga) sebanyak 11 RW dimana di tiap RW tersebut memiliki beberapa RT (Rukun Tetangga) dengan jumlah penduduk kurang lebih 2719 jiwa. Data statistik tahun 2019 dari badan pusat statistik Kabupaten Magelang (Hatta dkk., 2021). Dari data tersebut tercatat jumlah warga miskin (gakin) sebanyak 503 Kepala Keluarga (KK). Dari data warga miskin tersebut, penerima bantuan sosial 207 KK. Dengan adanya masalah tersebut, mengakibatkan adanya ketidak sinkronan penerima bantuan sosial. Masyarakat yang berhak menerima bisa jadi tidak menerima, sedangkan warga yang bukan merupakan warga miskin malah justru menerima bantuan sosial. Jika penerima bantuan sosial jelas terdata dengan baik, serta syarat kelayakan penerimaan bantuan

sosial yang terdata dengan baik, maka akan mengurangi ketidaksinkronan penerima bantuan sosial dan juga lebih memudahkan petugas di dalam pembagian bantuan sosial tersebut (Sugianto dkk., 2019).

Klasifikasi adalah proses penemuan model atau fungsi yang menggambarkan dan membedakan kelas data atau konsep yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui. Tujuan dari metode klasifikasi ini adalah untuk mengklasifikasikan nilai suatu variabel yang tidak diketahui dari variabel lain yang telah ditentukan sebelumnya (Lestari dkk., 2022). Fungsi klasifikasi memudahkan untuk mengidentifikasi kelayakan penerima dan bukan penerima bantuan PKM. Teknik klasifikasi adalah teknik pembelajaran yang digunakan untuk memprediksi nilai dari atribut kategori target (Sugianto dkk., 2019).

Salah satu metode klasifikasi yang bisa digunakan adalah *Random Forest*. *Random Forest* merupakan metode yang dapat meningkatkan hasil akurasi, karena dalam membangkitkan simpul anak untuk setiap *node* dilakukan secara acak. Metode ini digunakan untuk membangun pohon keputusan yang terdiri dari *root node*, *internal node*, dan *leaf node* dengan mengambil atribut dan data secara acak sesuai ketentuan yang diberlakukan (Prasojo & Haryatmi, 2021).

Berdasarkan uraian permasalahan diatas, dapat diketahui bahwa dalam mengatasi permasalahan ketidaksinkronan dalam penerima bantuan PKM pada Desa Alang Bon Bon. Oleh karena itu dalam penelitian ini akan berfokus terhadap algoritma *Random Forest* dalam melakukan klasifikasi terhadap masyarakat yang layak dan tidak layak menerima bantuan PKM. Dengan adanya klasifikasi data penduduk masyarakat agar penerima bantuan PKM bisa tepat sasaran.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diatas, rumusan masalah yang harus diselesaikan dalam penelitian ini yaitu bagaimana penerapan algoritma *Random Forest* dalam mengklasifikasikan penerima bantuan sosial pada tingkat akurasi di Desa Alang Bon – Bon Kabupaten Asahan.

1.3 Batasan Masalah

Klasifikasi ini memiliki cakupan yang luas, untuk itu agar penelitian lebih fokus, maka peneliti membuat batasan masalah yaitu :

1. Menggunakan algoritma *Random Forest* untuk klasifikasi kelayakan masyarakat untuk mendapatkan bantuan sosial.
2. Data yang digunakan dalam penelitian diambil dari data Observasi atau wawancara secara langsung di Desa Alang Bon-Bon.
3. Data yang digunakan adalah berjumlah 400 data penduduk Di Desa Alang Bon-Bon.
4. Melakukan pembagian *data training* 60% dan *testing* 40%, *data training* 70% dan *testing* 30%, serta *data training* 80% dan *testing* 20%.
5. Menggunakan *confusion matrix* sebagai pengujian terhadap hasil klasifikasi.
6. Atribut yang digunakan adalah nama kepala keluarga, pekerjaan, penghasilan, jumlah tanggungan, status rumah, kondisi atap, kondisi dinding, lansia, status lahan, serta ket.
7. Menggunakan bahasa pemrograman *Python*.

1.4 Tujuan Penelitian

Berdasarkan Rumusan masalah diatas, maka tujuan peneliti yang akan dicapai adalah:

1. Menentukan kelayakan keluarga penerima bantuan sosial berdasarkan hasil pengklasifikasian menggunakan *Random Forest*.
2. Mengukur seberapa lama waktu yang diperlukan dalam proses klasifikasi algoritma *Random Forest*

1.5 Manfaat Penelitian

Berdasarkan tujuan penelitian yang telah disebutkan di atas, maka hasil penelitian ini adalah:

1. Memberikan informasi dan masukan kepada pemerintah desa dalam menentukan kelayakan menerima bantuan sosial.

2. Memberi kontribusi ilmu pengetahuan kepada penulis khususnya dalam bidang *machine learning*.



BAB II TINJAUAN PUSTAKA

2.1 Kelayakan Bantuan Sosial

Definisi kelayakan Bantuan Sosial yang tertuang dalam Peraturan Presiden Republik Indonesia Nomor 7 Tahun 2005 tentang Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2004-2009 bahwa kemiskinan merupakan suatu kondisi dimana seseorang atau sekelompok orang, laki-laki dan perempuan, tidak terpenuhi hak-hak dasarnya untuk mempertahankan dan mengembangkan kehidupan yang bermartabat. Sedangkan definisi kemiskinan menurut Badan Pusat Statistik (BPS) dan Departemen Sosial, kemiskinan adalah ketidakmampuan individu dalam memenuhi kebutuhan dasar minimal untuk hidup layak (baik makanan maupun non makanan) (Almuqorobin, 2021).

Kemiskinan merupakan salah satu masalah yang dialami oleh beberapa negara berkembang, termasuk Indonesia. Banyak cara yang dilakukan untuk menanggulangi kemiskinan, diantaranya dengan program bantuan sosial untuk rakyat miskin (Abdurrahman & Ermawati, 2019). Berdasarkan penjelasan tersebut bisa disimpulkan bahwa kemiskinan adalah masalah yang dihadapi banyak negara berkembang dan menjadi pusat perhatian di negara manapun. Terdapat beberapa kriteria penelitian untuk mendapatkan bantuan sosial yaitu nama KRT, status pembangunan, status lahan, luas lantai, dinding, kondisi dinding, atap, kondisi atap, sumber air minum, cara memperoleh air minum, ibu hamil, anak sekolah, balita, lansia, penerima PKM.

2.2 Program Kelayakan Masyarakat (PKM)

PKM adalah salah satu program bantuan dari pemerintah yang melibatkan semua anggota keluarga, kepala rumah tangga harus memenuhi kualifikasi tertentu yang akan dipilih untuk mendapatkan bantuan tersebut, kualifikasi atau syarat untuk rumah tangga yang bisa mendapatkan bantuan tersebut dapat dibandingkan dengan melihat luas tanah sempit, jenis lantai bangunan, jenis dinding bangunan, WC, sumber air yang digunakan untuk sehari-hari, sumber penerangan dirumah, bahan

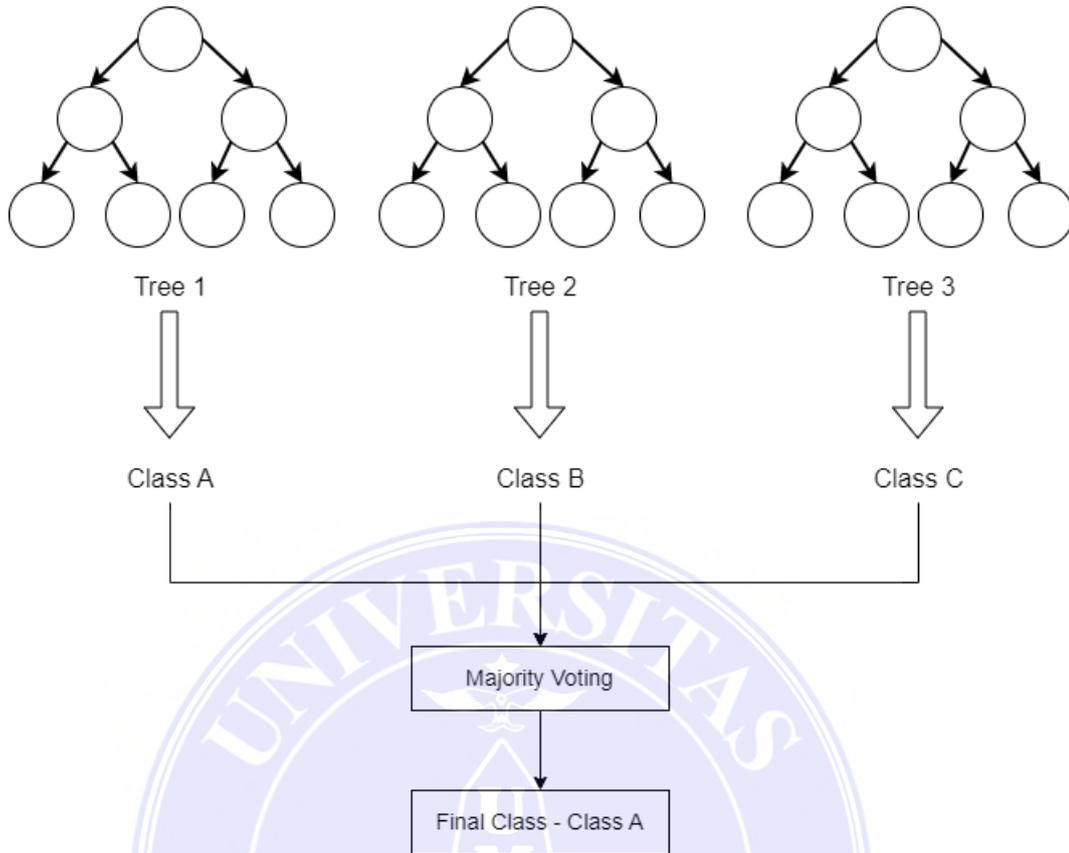
bakar yang digunakan untuk memasak, mampu atau tidak membayar anggota keluarga ke puskesmas, memiliki atau tidak ibu hamil atau anak balita, dan apakah memiliki anak yang sedang bersekolah. Dari beberapa kriteria tersebut rumah tangga yang berhak atau memenuhi, akan mendapat bantuan sesuai dengan anggaran biaya pemerintah pertahunnya. Dana tersebut akan diberikan 4kali dalam setahun, para penerima bantuan sosial ini berhak memenuhi sebuah kewajiban setiap bulannya bahwa dana tersebut digunakan untuk semestinya (Pamungkas dkk., 2019).

2.3 Klasifikasi

Klasifikasi dapat digambarkan sebagai berikut. *Data input*, disebut juga *training set*, terdiri atas banyak contoh (*record*), yang masing-masing memiliki beberapa atribut. Selanjutnya, tiap contoh diberi sebuah label *class* khusus. Tujuannya untuk menganalisa data *input* dan mengembangkan deskripsi atau model akurat untuk tiap *class* menggunakan fitur-fitur pada data. Deskripsi *class* ini digunakan untuk mengklasifikasikan data pengujian lainnya dengan label *class* tidak diketahui. Deskripsi tersebut juga dapat digunakan untuk memahami tiap *class* dalam data (Rizki dkk., 2020).

2.4 Random Forest

Random Forest merupakan salah satu metode CART (*Classification and Regression Tree*). Metode ini menggunakan konsep pohon keputusan (*decision tree*). Model ini dibentuk dari banyak pohon sehingga membentuk sebuah kumpulan pohon seperti hutan (*forest*) dengan menerapkan metode *bootstrap aggregating (bagging)* dan *random feature selection* (Iman & Wijayanto, 2021). Ilustrasi dari pohon keputusan dan pengambilan keputusan dalam metode *Random Forest* dapat dilihat dalam Gambar 2.1.



Gambar 2. 1 Ilustrasi Metode Random Forest

Sumber : (Iman & Wijayanto, 2021)

Menurut (Wuisan, 2020), berikut tahapan algoritma dalam pembuatan pohon keputusan dengan Algoritma *Random Forest*:

1. Perhatikan label pada data, jika sudah sama semua, maka akan dibentuk daun dengan nilai label data keseluruhan.
2. Menghitung nilai informasi dengan menggunakan semua data yang ada dengan formula:

$$Info(D) = - \sum_{i=1}^m p_i \log^2(p_i) \quad (2.1)$$

Dimana merupakan probabilitas tuple dalam D yang menjadi kelas dengan asumsi atau disebut juga *entropy* dari D merupakan rata-rata informasi yang diperlukan untuk identifikasi tuple dalam D.

Jika Nilai A adalah nilai diskrit maka data D akan dipisahkan sejumlah nilai data A sehingga nilai setiap cabang akan murni dan sejenis. Setelah percabangan pertama, jumlah percabangan yang mungkin terjadi diukur dengan persamaan:

$$infoA(D) = \sum_j^v \frac{|D_j|}{|D|} x InfoA(D) \quad (2.2)$$

3. Menghitung nilai informasi dengan formula
4. Untuk setiap atribut dengan memperhatikan isi data dari atribut. Dimana $\frac{|D_j|}{|D|}$ merupakan bobot dari partisij. $infoA(D)$ merupakan informasi yang diperlukan untuk mengklasifikasi *tuple* dari D pada partisi A. Semakin kecil hasil persamaan ini, semakin baik pula partisi yang dihasilkan. Nilai dari sebuah atribut menentukan penting tidaknya atribut tersebut dalam penyusunan pohon keputusan. Jika atribut bernilai kontinu, maka akan dicari *split_point* dengan cara mengurutkan seluruh data menurut atribut tersebut dari kecil ke besar, lalu di rata-rata antar satu data dengan data setelahnya. Nilai informasi akan dihitung menurut satu persatu calon *split_point* dan nilai *split_point* yang akan dipilih yang terkecil. Nilai *gain* untuk setiap atribut akan diperhitungkan dengan formula (2.3), nilai dengan gain tertinggi akan dijadikan cabang dalam pohon keputusan.

$$GAIN(A) = Info(D) - InfoA(D) \quad (2.3)$$

5. Setelah cabang pohon keputusan terbentuk, perhitungan dilakukan kembali seperti pada tahap 1 sampai 4. Namun jika cabang telah mencapai maksimal cabang yang diperbolehkan, daun akan terbentuk dengan nilai mayoritas dari nilai data.

2.5 Python

Pengertian *Python* (bahasa pemrograman) merupakan bahasa pemrograman tinggi yang bisa melakukan eksekusi sejumlah instruksi multi guna secara langsung

(interpretatif) dengan metode *Object Oriented Programming* dan juga menggunakan semantik dinamis untuk memberikan tingkat keterbacaan syntax. Sebagai bahasa pemrograman tinggi, *Python* dapat dipelajari dengan mudah karena telah dilengkapi dengan manajemen memori otomatis (Rahmadhika & Thantawi, 2021).

2.6 Confusion Matrix

Confusion matrix merupakan suatu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. *Confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem untuk diukur keakuratannya. *Confusion matrix* juga merupakan salah satu cara dalam melakukan visualisasi terhadap hasil pembelajaran sistem, visualisasi yang ditampilkan memuat dua kategori atau lebih (Almuqorobin, 2021).

Tabel 2. 1 *Confusion Matrix*

		Kelas Aktual	
		Positive	Negative
Kelas Prediksi	Positive	TP	FN
	Negative	FP	TN

Keterangan :

TP (True Positive) = jumlah data testing yang sebenarnya positif dan terprediksi positif.

TN (True Negative) = jumlah data testing yang sebenarnya negatif dan terprediksi negatif.

FP (False Positive) = jumlah data testing sebenarnya negatif dan terprediksi positif.

FN (False Negative) = jumlah data testing yang sebenarnya positif dan terprediksi negatif.

a. *Accuracy* akan menentukan seberapa baik model memprediksi dengan benar.

Persamaan *accuracy* yaitu:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.4)$$

- b. *Precision* merupakan akurasi antara data yang diminta dengan hasil yang diberikan oleh model. Persamaan *precision* yaitu:

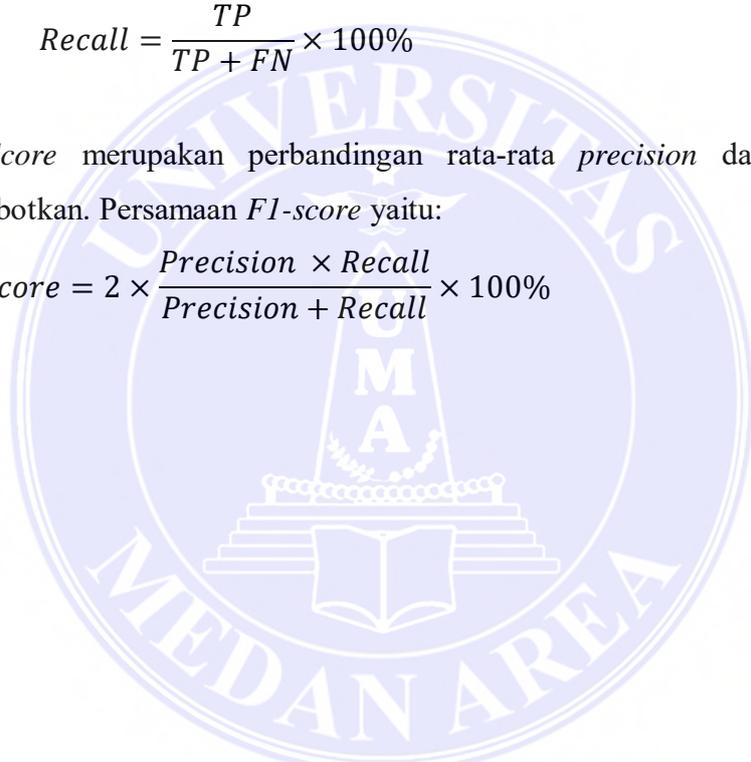
$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (2.5)$$

- c. *Recall* merupakan keberhasilan model dalam menemukan kembali sebuah informasi. Persamaan *recall* yaitu:

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (2.6)$$

- d. *F1-Score* merupakan perbandingan rata-rata *precision* dan *recall* yang dibobotkan. Persamaan *F1-score* yaitu:

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (2.7)$$



2.7 Penelitian Terdahulu

Tabel 2. 2 Penelitian Terdahulu

No	Penulis (Tahun)	Judul	Algoritma	Hasil
1.	(Rosid dkk., 2022)	Klasifikasi Penerima Bantuan Sosial Dengan Algoritma <i>Random Forest</i> Untuk Penanganan COVID 19	<i>Random Forest</i>	Penerima bantuan sosial untuk penanganan COVID19 dihasilkan klasifikasi <i>true</i> layak 31704 penerima bantuan sosial dan <i>true</i> tidak layak 2226 tidak layak menerima bantuan sosial. Tingkat akurasi klasifikasi penerima penerima bantuan sosial untuk penanganan COVID19 menggunakan algoritma <i>Random Forest performance accuracy</i> 93,44%.

2.	(Kurniawan dkk., 2023)	Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin.	<i>Random Forest</i>	Penerapan metode <i>Random Forest</i> dapat mengklasifikasikan data penerima bantuan raskin dengan benar dengan performa yang sangat baik yaitu <i>accuracy</i> sebesar 97,26%.
3.	(Iman & Wijayanto, 2021)	Klasifikasi Rumah Tangga Penerima Beras Miskin (Raskin)/Beras Sejahtera (Rastra) di Provinsi Jawa Barat Tahun 2017 dengan Metode <i>Random Forest</i> dan <i>Support Vector Machine</i> .	<i>Random Forest</i> dan <i>Support Vector Machine</i> .	Data yang digunakan bersumber dari data mikro Susenas Provinsi Jawa Barat tahun 2017. Data berjumlah 23.756 yang merupakan seluruh rumah tangga di Jawa Barat. Rata-rata akurasi yang dihasilkan dengan metode <i>Random Forest</i> adalah sebesar 71–72% dan metode SVM adalah sebesar 72%. Hal ini menunjukkan model yang dibentuk dari kedua metode sudah

				baik dalam melakukan pengelompokan (klasifikasi) rumah tangga penerima program bantuan Raskin/Rastra. Secara keseluruhan, selisih kinerja klasifikasi kedua metode random forest dan SVM tidak signifikan. Maka, kedua metode klasifikasi memiliki kinerja yang setara dan sama baiknya dalam melakukan klasifikasi.
4.	(Devella dkk., 2020)	Implementasi <i>Random Forest</i> Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT	<i>Random Forest</i>	Data latih dan data uji yang digunakan masing-masing sebanyak 100 dan 15 untuk setiap motif Songket Palembang. Hasil pengujian menunjukkan bahwa metode SIFT dan Random Forest untuk klasifikasi citra motif kain Songket

				Palembang dapat memberikan akurasi yang cukup baik, dimana metode SIFT dan <i>Random Forest</i> mampu menghasilkan rata-rata <i>overall accuracy</i> 92,98%, per class <i>accuracy</i> 94,07%, <i>presision</i> 92,98%, dan <i>recall</i> 89,74%.
5.	(Yusuf dkk., 2020)	Implementasi Algoritma <i>Naive Bayes</i> Dan <i>Random Forest</i> Dalam Memprediksi Prestasi Akademik Mahasiswa Universitas Islam Negeri Ar-Raniry Banda Aceh	<i>Naive Bayes</i> Dan <i>Random Forest</i>	Persentase <i>data training</i> yang digunakan sebesar 60%, 70%, 80% dan 90% dari 1500 sampel, sampel yang tersisa dari besaran persentase pada tiap-tiap <i>data training</i> digunakan sebagai <i>data testing</i> . Dari pengujian <i>cross-validation</i> pada kedua algoritma diperoleh hasil algoritma <i>Random Forest</i> lebih unggul dibandingkan

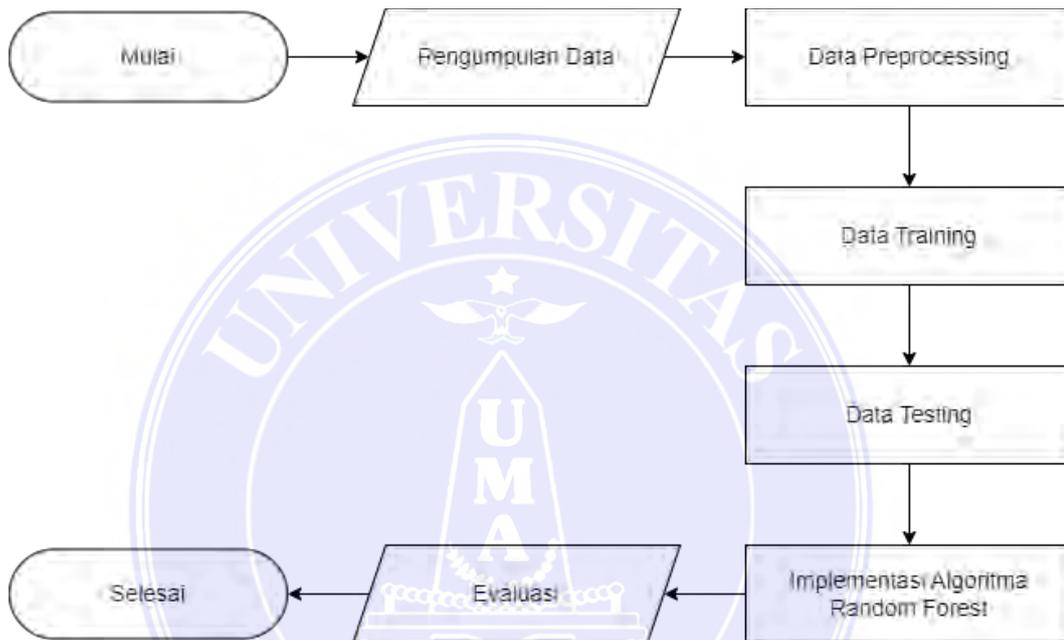
				<p><i>Naïve Bayes</i>. Sedangkan pada pengujian <i>percentage split</i>, algoritma <i>Naïve Bayes</i> lebih unggul dengan tingkat keakuratannya sebesar 78,0% sedangkan <i>Random Forest</i> sebesar 76,6%. Artinya algoritma <i>Naive Bayes</i> lebih baik tingkat keakuratannya.</p>
6.	(Bawono & Wasono, 2019)	Perbandingan Metode <i>Random Forest</i> dan <i>Naive Bayes</i> Untuk Klasifikasi Debitur Berdasarkan Kualitas Kredit.	<i>Random Forest</i> dan <i>Naive Bayes</i>	Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari salah satu Bank. Data yang memiliki 10 atribut yang terdiri dari Tanggal Akad, Tanggal Jatuh Tempo, Plafond, Saldo September 2017, Tunggakan Pokok, Tunggakan Bunga, Suku Bunga, Jaminan, Nilai Jaminan, Jangka Waktu (bulan) dan

				<p>satu label yakni Tingkat Kolektibilitas.</p> <p>Jumlah keseluruhan dataset yang digunakan ada 6320 data. Hasil yang didapat <i>Random Forest</i> memiliki hasil akurasi tertinggi yakni sebesar 98,16% disusul <i>Naïve Bayes</i> 95,93%. Sehingga dengan <i>Random Forest</i> nasabah yang memiliki kredit bermasalah hanya 3,22%</p>
7.	(Zailani & Hanun, 2020)	<p>Penerapan Algoritma <i>Random Forest</i></p> <p>Klasifikasi <i>Random Forest</i></p> <p>Untuk Penentuan Kelayakan</p> <p>Pemberian Kredit Di Koperasi</p> <p>Mitra Sejahtera</p>	<i>Random Forest</i>	<p>Hasil pengujian dengan algoritma klasifikasi <i>Random Forest</i> mampu menganalisis kredit yang bermasalah dan yang debitur yang tidak bermasalah dengan nilai akurasi sebesar 87,88%</p>

BAB III METODOLOGI PENELITIAN

3.1 Prosedur Penelitian

Prosedur yang dilakukan dalam penelitian ini dimulai dengan menentukan alur kerja *Random Forest* untuk mendapatkan hasil yang akurat. Alur penelitian bisa dilihat pada Gambar 3.1.



Gambar 3. 1 Prosedur Penelitian

Uraian prosedur penelitian yang digambarkan pada Gambar 3.1 yaitu pengumpulan data yang berfungsi untuk mengumpulkan data yang di ambil langsung dari Kantor Kelurahan Desa Alang Bon – Bon Kecamatan Aek Kuasan Kabupaten Asahan. Selanjutnya, *data preprocessing* yang berfungsi sebagai pengecekan data yang kosong, data yang terduplikat, dan lain sebagainya. Kemudian tahapan pembagian *data training* dan *data testing* yaitu untuk membagi data mana yang akan dilakukan sebagai latih dan data mana yang dilakukan sebagai penguji. Selanjutnya, melakukan implementasi terhadap algoritma *machine learning* yaitu *Random Forest*. Terakhir adalah evaluasi dengan menggunakan *confusion matrix*, fungsinya adalah menguji

perhitungan akurasi pada klasifikasi bantuan sosial di Kantor Kelurahan Desa Alang Bon – Bon Kecamatan Aek Kuasan Kabupaten Asahan.

3.2 Data dan Sumber Data

Penelitian ini menggunakan data primer, yaitu data penerima bantuan sosial di Kantor Kelurahan Desa Alang Bon – Bon Kecamatan Aek Kuasan Kabupaten Asahan sejumlah 400 data yang mempunyai atribut nama kepala keluarga, pekerjaan, penghasilan, jumlah tanggungan, status rumah, kondisi atap, kondisi dinding, lansia, status lahan, serta ket. Adapun pengumpulan data dilakukan dengan survei, yaitu teknik pengumpulan data atau informasi pada populasi yang besar dengan menggunakan sampel yang relatif lebih kecil. Metode ini juga dilakukan dengan mengadakan pengamatan secara langsung terhadap suatu proses yang tengah berjalan atau berlangsung.

3.3 Teknik Analisis Data

Teknik analisa data adalah sebuah kegiatan berupa menganalisa suatu penelitian dengan mengontrol seluruh data yang berasal dari instrumen penelitian, seperti dokumen, catatan, rekaman dan lain-lain. Tujuannya adalah untuk memahami data dan memperoleh kesimpulan dengan lebih mudah. Teknik analisis data yang digunakan pada penelitian adalah sebagai berikut:

1. Menganalisis data secara menyeluruh untuk mengetahui gambaran umum mengenai data bantuan sosial di Kantor Kelurahan Desa Alang Bon – Bon Kecamatan Aek Kuasan Kabupaten Asahan, serta menentukan variabel apa saja yang digunakan pada penelitian ini.
2. Membagi data menjadi 2 bagian yaitu *data training* dan *data testing*. Adapun prosedur yang dilakukan pada penelitian ini memiliki 3 tahapan dengan jumlah proporsi yang berbeda, dapat dilihat pada Tabel 3.1.

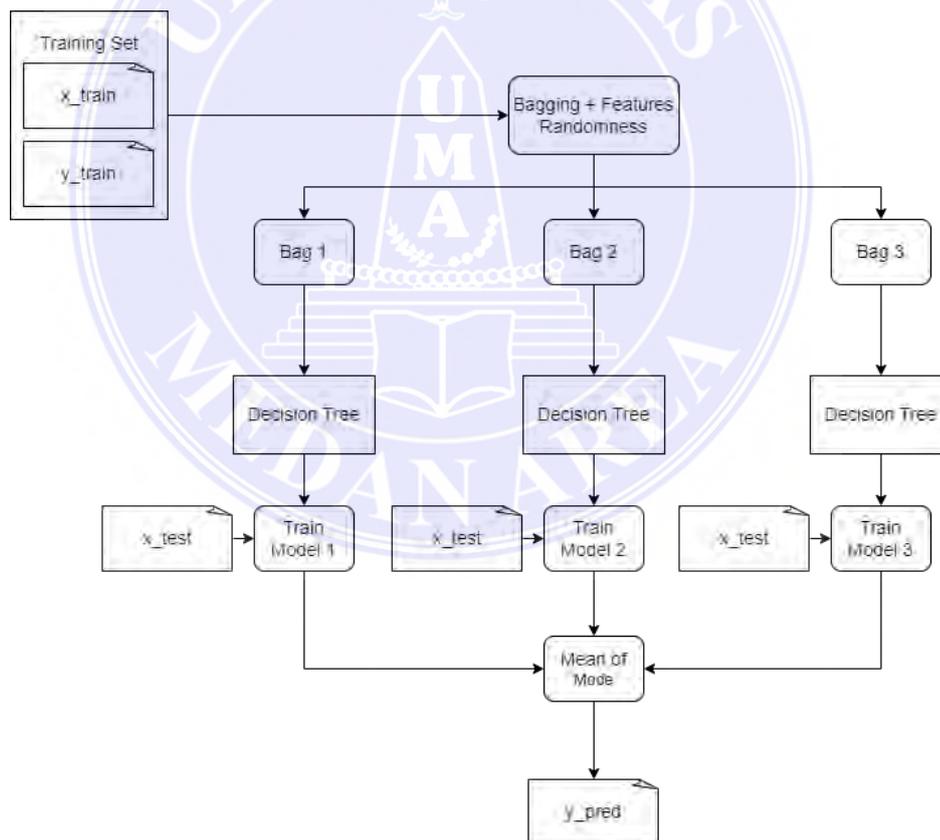
Tabel 3. 1 Proporsi Data

Prosedur	<i>Data Training</i>	<i>Data Testing</i>
1	60 %	40 %
2	70 %	30 %
3	80 %	20 %

Alasan peneliti menggunakan proporsi data yang berbeda – beda ialah ingin mengetahui apakah akan ada pengaruh hasil akurasi dalam pengklasifikasian terhadap bantuan sosial di Kantor Kelurahan Desa Alang Bon – Bon Kecamatan Aek Kuasan Kabupaten Asahan dengan menggunakan algoritma *Random Forest*.

3. Membuat model klasifikasi dari data train menggunakan algoritma *Random Forest*.
4. Menghitung ketepatan klasifikasi dengan cara menghitung akurasi dalam proses pengklasifikasian.

3.4 Implementasi *Random Forest*



Gambar 3. 2 Implementasi *Random Forest*

Pada Gambar 3.2 dijabarkan bahwa *Random Forest* dapat dibangun menggunakan bagging dengan pemilihan atribut acak. Pada setiap bag di latih menggunakan model yang sama yaitu *Decision Tree* sehingga menghasilkan model terlatih dengan hasil yang berbeda beda dikarenakan dataset yang berbeda walau menggunakan model yang sama. Selanjutnya prediksi tiap model akan di satukan menggunakan *majority voting* untuk mendapatkan nilai prediksi (y_{pred}) sebagai hasil klasifikasi.

No	pekerjaan	penghasilan	jum tanggungan	status rumah	kondisi atap	kondisi dinding	Lansia	status lahan	ket
1	IBU RUMAH TANGGA	1,000,000	2	MILIK SENDIRI	SENG	papan	3	TIDAK ADA	LAYAK
2	PETANI	1,500,000	1	MILIK SENDIRI	GENTENG	BETON	0	ADA	TIDAK LAYAK
3	BURUH TANI	500,000	3	MILIK SENDIRI	SENG	papan	2	TIDAK ADA	LAYAK
4	BURUH TANI	500,000	5	MILIK SENDIRI	SENG	BETON	0	TIDAK ADA	LAYAK
5	IBU RUMAH TANGGA	1,000,000	3	MILIK SENDIRI	SENG	papan	3	TIDAK ADA	LAYAK
6	IBU RUMAH TANGGA	2,000,000	2	MILIK SENDIRI	GENTENG	BETON	0	ADA	TIDAK LAYAK
7	BURUH TANI	500,000	0	NUMPANG	SENG	papan	1	TIDAK ADA	LAYAK
8	BURUH TANI	500,000	0	NUMPANG	SENG	papan	0	TIDAK ADA	LAYAK
9	IBU RUMAH TANGGA	1,000,000	3	MILIK SENDIRI	SENG	BETON	1	ADA	LAYAK
10	IBU RUMAH TANGGA	1,000,000	3	MILIK SENDIRI	SENG	papan	0	TIDAK ADA	LAYAK
11	kedai	1,500,000	2	MILIK SENDIRI	GENTENG	BETON	1	ADA	TIDAK LAYAK
12	PETANI	1,000,000	5	MILIK SENDIRI	SENG	papan	0	TIDAK ADA	LAYAK
13	kedai	1,000,000	1	MILIK SENDIRI	SENG	BETON	0	ADA	TIDAK LAYAK
14	kedai	1,000,000	1	MILIK SENDIRI	SENG	BETON	0	ADA	TIDAK LAYAK
15	kedai	1,000,000	2	MILIK SENDIRI	SENG	papan	3	TIDAK ADA	LAYAK
16	PETANI	1,500,000	3	MILIK SENDIRI	GENTENG	BETON	1	ADA	TIDAK LAYAK

Gambar 3. 3 Data Latih

Gambar 3.3 merupakan data latih yang berjumlah 16 dan akan dilakukan perhitungan manual menggunakan algoritma *Random Forest*, adapun atribut yang digunakan yaitu pekerjaan, penghasilan, jumlah tanggungan, status rumah, kondisi atap, kondisi dinding, lansia, status lahan, serta ket.

Tabel 3. 2 Node 1 Entropy Dan Gain

	Keterangan	Jumlah	Ket (Layak)	Ket (Tidak Layak)	Entropy	Gain
Total		16	10	6	0,954	
Pekerjaan	IBU RUMAH TANGGA	5	4	1	0,721	0,353
	PETANI	3	1	2	0,918	
	BURUH TANI	4	4	0	0	
	kedai	4	1	3	0,811	

Penghasilan	< 900000	4	4	0	0	0,2 04
	> 900000	12	6	6	1	
Jum Tanggungan	< 4	14	8	6	0,985	0,0 92
	> 4	2	2	0	0	
status rumah	MILIK SENDIRI	14	8	6	0,985	0,0 92
	NUMPANG	2	2	0	0	
kondisi atap	SENG	12	10	2	0,650	0,4 66
	GENTENG	4	0	4	0	
kondisi dinding	papan	8	8	0	0	0,5 48
	BETON	8	2	6	0,811	
Lansia	< 2	12	6	6	1	0,2 04
	>= 2	4	4	0	0	
status lahan	TIDAK ADA	9	9	0	0	0,6 95
	ADA	7	1	6	0,591	

Pada Tabel 3.2 merupakan perhitungan *entropy* total dilakukan dengan menggunakan persamaan (2.1 dan 2.2). Rumus tersebut digunakan untuk mengukur ketidakpastian dalam himpunan data. Adapun perhitungan *entropy* pada node 1 adalah sebagai berikut :

$$Entropy(Total) = \left(-\frac{10}{16} * \log_2\left(\frac{10}{16}\right)\right) + \left(-\frac{6}{16} * \log_2\left(\frac{6}{16}\right)\right) = 0.954$$

$$Entropy(Ibu Rumah Tangga) = \left(-\frac{4}{5} * \log_2\left(\frac{4}{5}\right)\right) + \left(-\frac{1}{5} * \log_2\left(\frac{1}{5}\right)\right) = 0.721$$

$$Entropy(Petani) = \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{1} * \log_2\left(\frac{2}{1}\right)\right) = 0.918$$

$$Entropy(Buruh Tani) = \left(-\frac{4}{4} * \log_2\left(\frac{4}{4}\right)\right) + \left(-\frac{0}{4} * \log_2\left(\frac{0}{4}\right)\right) = 0$$

$$Entropy(Kedai) = \left(-\frac{1}{4} * \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} * \log_2\left(\frac{3}{4}\right)\right) = 0.811$$

$$Entropy(< 900000) = \left(-\frac{4}{4} * \log_2\left(\frac{4}{4}\right)\right) + \left(-\frac{0}{4} * \log_2\left(\frac{0}{4}\right)\right) = 0$$

$$Entropy(> 900000) = \left(-\frac{6}{12} * \log_2\left(\frac{6}{12}\right)\right) + \left(-\frac{6}{12} * \log_2\left(\frac{6}{12}\right)\right) = 1$$

$$Entropy(< 4) = \left(-\frac{8}{14} * \log_2\left(\frac{8}{14}\right)\right) + \left(-\frac{6}{14} * \log_2\left(\frac{6}{14}\right)\right) = 0.985$$

$$Entropy(> 4) = \left(-\frac{2}{2} * \log_2\left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) = 0$$

$$Entropy(Milik Sendiri) = \left(-\frac{8}{14} * \log_2\left(\frac{8}{14}\right)\right) + \left(-\frac{6}{14} * \log_2\left(\frac{6}{14}\right)\right) = 0.985$$

$$Entropy(Numpang) = \left(-\frac{2}{2} * \log_2\left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) = 0$$

$$Entropy(Seng) = \left(-\frac{10}{12} * \log_2\left(\frac{10}{12}\right)\right) + \left(-\frac{2}{12} * \log_2\left(\frac{2}{12}\right)\right) = 0.65$$

$$Entropy(Genteng) = \left(-\frac{0}{4} * \log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} * \log_2\left(\frac{4}{4}\right)\right) = 0$$

$$Entropy(Papan) = \left(-\frac{8}{8} * \log_2\left(\frac{8}{8}\right)\right) + \left(-\frac{0}{8} * \log_2\left(\frac{0}{8}\right)\right) = 0$$

$$Entropy(Beton) = \left(-\frac{2}{8} * \log_2\left(\frac{2}{8}\right)\right) + \left(-\frac{6}{8} * \log_2\left(\frac{6}{8}\right)\right) = 0.81$$

$$Entropy(< 2) = \left(-\frac{6}{12} * \log_2\left(\frac{6}{12}\right)\right) + \left(-\frac{6}{12} * \log_2\left(\frac{6}{12}\right)\right) = 1$$

$$Entropy(\geq 2) = \left(-\frac{4}{4} * \log_2\left(\frac{4}{4}\right)\right) + \left(-\frac{0}{4} * \log_2\left(\frac{0}{4}\right)\right) = 0$$

$$Entropy(Tidak Ada) = \left(-\frac{9}{9} * \log_2\left(\frac{9}{9}\right)\right) + \left(-\frac{0}{9} * \log_2\left(\frac{0}{9}\right)\right) = 0$$

$$Entropy(Ada) = \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) = 0.59$$

Adapun perhitungan *Gain* pada node 1 adalah sebagai berikut :

Gain(Pekerjaan)

$$= 0.954 - \left(\left(\frac{5}{16} \right) * 0.721 \right) + \left(\left(\frac{3}{16} \right) * 0.918 \right) + \left(\left(\frac{4}{16} \right) * 0 \right) + \left(\left(\frac{4}{16} \right) * 0.811 \right) = 0.353$$

$$Gain(Penghasilan) = 0.954 - \left(\left(\frac{4}{16} \right) * 0 \right) + \left(\left(\frac{12}{16} \right) * 1 \right) = 0.204$$

$$Gain(Jumlah Tanggungan) = 0.954 - \left(\left(\frac{14}{16} \right) * 0.985 \right) + \left(\left(\frac{2}{16} \right) * 0 \right) = 0.092$$

$$Gain(Status Rumah) = 0.954 - \left(\left(\frac{14}{16} \right) * 0.985 \right) + \left(\left(\frac{2}{16} \right) * 0 \right) = 0.092$$

$$Gain(Kondisi Atap) = 0.954 - \left(\left(\frac{12}{16} \right) * 0.650 \right) + \left(\left(\frac{4}{16} \right) * 0 \right) = 0.466$$

$$Gain(Kondisi Dinding) = 0.954 - \left(\left(\frac{8}{16} \right) * 0 \right) + \left(\left(\frac{8}{16} \right) * 0.811 \right) = 0.548$$

$$Gain(Lansia) = 0.954 - \left(\left(\frac{12}{16} \right) * 1 \right) + \left(\left(\frac{4}{16} \right) * 0 \right) = 0.204$$

$$Gain(Status Lahan) = 0.954 - \left(\left(\frac{9}{16} \right) * 0 \right) + \left(\left(\frac{7}{16} \right) * 0.591 \right) = 0.695$$

Tabel 3. 3 Node 2 Entropy Dan Gain

	Keterangan	Jumlah	Ket (Layak)	Ket (Tidak Layak)	Entropy	Gain
Total		7	1	6	0,591	
Pekerjaan	IBU RUMAH TANGGA	2	1	1	1	0,305
	PETANI	2	0	2	0	
	BURUH TANI	0	0	0	0	
	kedai	3	0	3	0	
Penghasilan	< 900000	0	0	0	0	0
	> 900000	7	1	6	0,591	

Jum Tanggungan	< 4	7	1	6	0,591	0
	> 4	0	0	0	0	
status rumah	MILIK SENDIRI	7	1	6	0,591	0
	NUMPANG	0	0	0	0	
kondisi atap	SENG	3	1	2	0,918	0,198
	GENTENG	4	0	4	0	
kondisi dinding	papan	0	0	0	0	0
	BETON	7	1	6	0,591	
lansia	< 2	7	1	6	0,591	0
	>= 2	0	0	0	0	

Pada Tabel 3.3 merupakan perhitungan *entropy* total dilakukan dengan menggunakan persamaan (2.1 dan 2.2). Rumus tersebut digunakan untuk mengukur ketidakpastian dalam himpunan data. Adapun perhitungan *entropy* pada node 2 adalah sebagai berikut :

$$Entropy(Total) = \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) = 0.591$$

$$Entropy(Ibu Rumah Tangga) = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$Entropy(Petani) = \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} * \log_2\left(\frac{2}{2}\right)\right) = 0$$

$$Entropy(Buruh Tani) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(Kedai) = \left(-\frac{0}{3} * \log_2\left(\frac{0}{3}\right)\right) + \left(-\frac{3}{3} * \log_2\left(\frac{3}{3}\right)\right) = 0$$

$$Entropy(< 900000) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(> 900000) = \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) = 0.591$$

$$Entropy(< 4) = \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) = 0.591$$

$$Entropy(> 4) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(Milik Sendiri) = \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) = 0.591$$

$$Entropy(Numpang) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(Seng) = \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) = 0.918$$

$$Entropy(Genteng) = \left(-\frac{0}{4} * \log_2\left(\frac{0}{4}\right)\right) + \left(-\frac{4}{4} * \log_2\left(\frac{4}{4}\right)\right) = 0$$

$$Entropy(Papan) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(Beton) = \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) = 0.591$$

$$Entropy(< 2) = \left(-\frac{1}{7} * \log_2\left(\frac{1}{7}\right)\right) + \left(-\frac{6}{7} * \log_2\left(\frac{6}{7}\right)\right) = 0.591$$

$$Entropy(\geq 2) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

Adapun perhitungan *Gain* pada node 2 adalah sebagai berikut :

$$\begin{aligned} Gain(Pekerjaan) &= 0.591 - \left(\left(\frac{2}{7}\right) * 1\right) + \left(\left(\frac{2}{7}\right) * 0\right) + \left(\left(\frac{0}{7}\right) * 0\right) + \left(\left(\frac{3}{7}\right) * 0\right) \\ &= 0.305 \end{aligned}$$

$$Gain(Penghasilan) = 0.591 - \left(\left(\frac{0}{7}\right) * 0\right) + \left(\left(\frac{7}{7}\right) * 0.591\right) = 0$$

$$Gain(Jumlah Tanggungan) = 0.591 - \left(\left(\frac{7}{7}\right) * 0.591\right) + \left(\left(\frac{0}{7}\right) * 0\right) = 0$$

$$Gain(\text{Status Rumah}) = 0.591 - \left(\left(\frac{7}{7} \right) * 0.591 \right) + \left(\left(\frac{0}{7} \right) * 0 \right) = 0$$

$$Gain(\text{Kondisi Atap}) = 0.591 - \left(\left(\frac{3}{7} \right) * 0.918 \right) + \left(\left(\frac{4}{7} \right) * 0 \right) = 0.198$$

$$Gain(\text{Kondisi Dinding}) = 0.591 - \left(\left(\frac{0}{7} \right) * 0 \right) + \left(\left(\frac{7}{7} \right) * 0.591 \right) = 0$$

$$Gain(\text{Lansia}) = 0.591 - \left(\left(\frac{7}{7} \right) * 0.591 \right) + \left(\left(\frac{0}{7} \right) * 0 \right) = 0$$

Tabel 3. 4 Node 3 Entropy Dan Gain

	Keterangan	Jumlah	Ket (Layak)	Ket (Tidak Layak)	Entropy	Gain
Total		2	1	1	1	
Penghasilan	< 900000	0	0	0	0	-0,408
	> 900000	2	1	1	1	
Jumlah Tanggungan	< 4	2	1	1	1	-0,408
	> 4	0	0	0	0	
status rumah	MILIK SENDIRI	2	1	1	1	-0,408
	NUMPANG	0	0	0	0	
kondisi atap	SENG	1	1	0	0	0,591
	GENTENG	1	0	1	0	
kondisi dinding	papan	0	0	0	0	-0,408
	BETON	2	1	1	1	

						-
lansia	< 2	2	1	1	1	0,40
	>= 2	0	0	0	0	8

Pada Tabel 3.4 merupakan perhitungan *entropy* total dilakukan dengan menggunakan persamaan (2.1 dan 2.2). Rumus tersebut digunakan untuk mengukur ketidakpastian dalam himpunan data. Adapun perhitungan *entropy* pada node 3 adalah sebagai berikut :

$$Entropy(Total) = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$Entropy(< 900000) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(> 900000) = \left(-\frac{1}{27} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$Entropy(< 4) = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$Entropy(> 4) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(Milik Sendiri) = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$Entropy(Numpang) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(Seng) = \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) = 0$$

$$Entropy(Genteng) = \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) = 0$$

$$Entropy(Papan) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(Beton) = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$Entropy(< 2) = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1$$

$$Entropy(\geq 2) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

Adapun perhitungan *Gain* pada node 3 adalah sebagai berikut :

$$Gain(Penghasilan) = 1 - \left(\left(\frac{0}{2}\right) * 0\right) + \left(\left(\frac{2}{2}\right) * 1\right) = -0.408$$

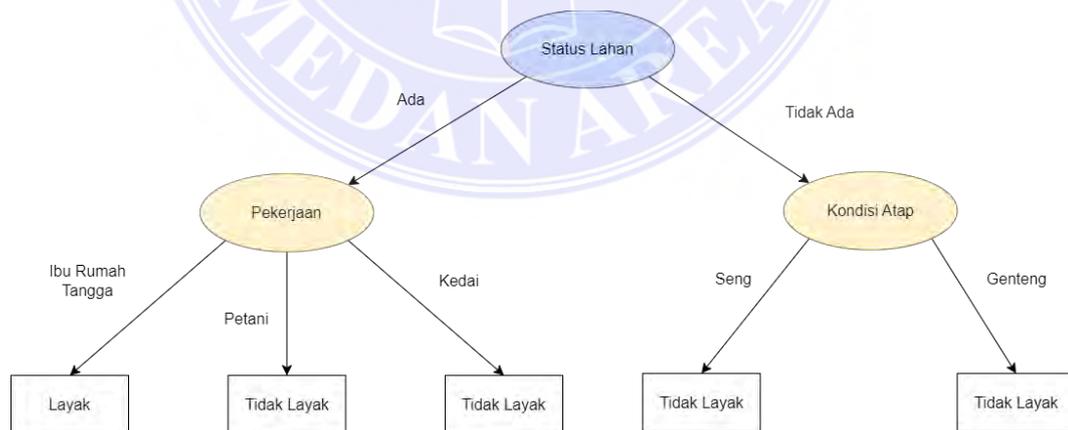
$$Gain(Jumlah Tanggungan) = 1 - \left(\left(\frac{2}{2}\right) * 1\right) + \left(\left(\frac{0}{2}\right) * 0\right) = -0.408$$

$$Gain(Status Rumah) = 1 - \left(\left(\frac{2}{2}\right) * 1\right) + \left(\left(\frac{0}{2}\right) * 0\right) = -0.408$$

$$Gain(Kondisi Atap) = 1 - \left(\left(\frac{1}{2}\right) * 0\right) + \left(\left(\frac{1}{2}\right) * 0\right) = 0.591$$

$$Gain(Kondisi Dinding) = 1 - \left(\left(\frac{0}{2}\right) * 0\right) + \left(\left(\frac{2}{2}\right) * 1\right) = -0.408$$

$$Gain(Lansia) = 1 - \left(\left(\frac{2}{2}\right) * 1\right) + \left(\left(\frac{0}{2}\right) * 0\right) = -0.408$$



Gambar 3. 4 Pohon Keputusan

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Dari hasil yang sudah dilakukan pada klasifikasi terhadap kelayakan masyarakat untuk mendapatkan bantuan sosial dapat disimpulkan bahwa hasil prosedur ke-3 dengan menggunakan *data training* 80% dan *data testing* 20% mendapatkan akurasi sebesar 97%, *precision* 97%, *recall* 97%, dan *f1-score* 97%. Ini dapat disimpulkan bahwa pada prosedur ke-3 adalah model terbaik karena data latih yang digunakan lebih banyak digunakan daripada prosedur sebelumnya, sehingga mendapatkan performa yang tinggi serta prediksinya yang akurat. Kemudian pada prosedur tersebut terdapat 2 kesalahan prediksi yaitu pada kepala keluarga Gustina Br Panjaitan dan Asma.

5.2 Saran

Adapun saran yang dapat dikembangkan pada penelitian selanjutnya adalah:

1. Permasalahan klasifikasi tidak begitu akurat dikarenakan data sangat sedikit sekali. Sehingga penelitian selanjutnya dapat menambah jumlah data yang sangat banyak serta menambah jumlah atributnya.
2. Penelitian selanjutnya dapat membuat aplikasi *website* dengan menggunakan *framework flask*.
3. Menggunakan metode lain seperti *C.5*, *Adaboost*, *Naïve Bayes*, dan *Logistic Regression*.

DAFTAR PUSTAKA

- Abdurrahman, M. A., & Ermawati, W. J. (2019). Pengaruh Leverage, Financial Distress dan Profitabilitas terhadap Konservatisme Akuntansi pada Perusahaan Pertambangan di Indonesia Tahun 2013-2017. *Jurnal Manajemen dan Organisasi*, 9(3), 164–173. <https://doi.org/10.29244/jmo.v9i3.28227>
- Almuqorobin, A. (2021). *Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan Menggunakan Random Forest (Studi Kasus: Kelurahan Desa Balerejo)*. Universitas Muhammadiyah Megelang.
- Bawono, B., & Wasono, R. (2019). Perbandingan Metode Random Forest dan Naive Bayes Untuk Klasifikasi Debitur Berdasarkan Kualitas Kredit. *Jurnal Sains dan Sistem Informasi*, 3(7), 343–348. <http://prosiding.unimus.ac.id>
- Devella, S. ... Rahmawati, F. N. (2020). Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, 7(2), 310–320. <https://doi.org/10.35957/jatisi.v7i2.289>
- Hatta, H. R. ... Kridalaksana, A. H. (2021). *Pengembangan Game Adventure of Utan Menggunakan Challenge Cadence Skill-Theme*. 16(1), 1–6.
- Hermansyah, D. (2022). Penentuan Status Mutu Air Sungai Kapuas Menggunakan Metode Storet Dan Logika Fuzzy Mamdani. *Prisma Fisika*, 10(2), 128–134.
- Iman, Q., & Wijayanto, A. W. (2021). Klasifikasi Rumah Tangga Penerima Beras Miskin (Raskin)/Beras Sejahtera (Rastra) di Provinsi Jawa Barat Tahun 2017 dengan Metode Random Forest dan Support Vector Machine. *Jurnal Sistem dan Teknologi Informasi (Justin)*, 9(2), 178. <https://doi.org/10.26418/justin.v9i2.44137>
- Kurniawan, I. ... Korespondensi, P. (2023). *Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin*. 10(2), 421–428. <https://doi.org/10.25126/jtiik.202396225>
- Lestari, C. ... Siahaan, S. (2022). *Analisis Sentimen Pandangan Netizen Indonesia*

Terhadap Vaksin COVID-19 Menggunakan Multi-Layer Perceptron. 9(4), 2795–2803.

Pamungkas, V. C. ... Wihandika, R. C. (2019). Klasifikasi Penerimaan Program Keluarga Harapan (PKH) Menggunakan Metode Learning Vector Quantization (Studi Kasus Desa Kedungjati). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3), 2659–2666.

Prasojo, B., & Haryatmi, E. (2021). *Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random. 02, 79–89.*

Pratama, Y. C. (2022). Analisis Faktor-Faktor Yang Mempengaruhi Kemiskinan Di Indonesia. *Esensi*, 4(2), 45–53. <https://doi.org/10.15408/ess.v4i2.1966>

Rahmadhika, M. K., & Thantawi, A. M. (2021). Rancang Bangun Aplikasi Face Recognition Pada Pendekatan CRM Menggunakan Opencv Dan Algoritma Haarcascade. *IKRA-ITH INFORMATIKA: Jurnal Komputer dan Informatika*, 5(1), 109–118.

Rizki, M. Y. ... Pembahasan, A. D. A. N. (2020). Implementasi C45 Dalam Memprediksi Index Prestasi Mahasiswa / i Menurut Kebiasaan Belajar. *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)*, 297–303.

Rosid, A. ... Dwilestari, G. (2022). *Klasifikasi Penerima Bantuan Sosial Dengan Algoritma Random Forest Untuk Penanganan COVID 19. 10.*

Sugianto, C. A. ... Mining, D. (2019). *Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama). 18(4), 321–331.*

Wuisan, D. S. S. (2020). Analisis Sistem Informasi Pengelolaan Data Transaksi Pembelian Consumer Untuk Manajemen Pembelian Perusahaan. *Infotech: Journal of Technology Information*, 6(1), 29–34. <https://doi.org/10.37365/jti.v6i1.75>

Yusuf, B. ... Ellyadi, M. (2020). Implementasi Algoritma Naive Bayes Dan Random Forest Dalam Memprediksi Prestasi Akademik Mahasiswa Universitas Islam Negeri Ar-Raniry Banda Aceh. *Cyberspace: Jurnal*

Pendidikan Teknologi Informasi, 4(1), 50.
<https://doi.org/10.22373/cj.v4i1.7247>

Zailani, A. U., & Hanun, N. L. (2020). Penerapan Algoritma Klasifikasi Random Forest Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera. *Infotech: Journal of Technology Information*, 6(1), 7–14.
<https://doi.org/10.37365/jti.v6i1.61>



LAMPIRAN PROGRAM

```
1. import pandas as pd
2.
3. dataframe = pd.read_excel('/content/datas.xlsx')
4.
5. import matplotlib.pyplot as plt
6. # Membuat jumlah visualisasi dari masing-masing klasifikasi
7. data = dataframe.ket.value_counts()
8. data.plot(kind='pie', autopct='%0.2f%%')
9. plt.show()
10.
11. # Label Encoder
12. from sklearn.preprocessing import LabelEncoder
13.
14. # Menyalin / copy dataframe agar dataframe awal tetap utuh
15. df = dataframe.copy()
16.
17. label = LabelEncoder()
18.
19. # Membuat list dari nama kolom data kategori
20. categorical_data = ['pekerjaan', 'status_rumah', 'kondisi_atap', 'kondisi_dinding',
    'status_lahan', 'ket']
21.
22. # Mengubah setiap data kategori menjadi numerik dengan encoder
23. for kolom in categorical_data:
24.     df[kolom] = label.fit_transform(df[kolom])
25.
26. # Menentukan X & y
27. kolom = ['pekerjaan', 'penghasilan', 'jum tanggungan', 'status rumah', 'kondisi
    atap', 'kondisi dinding', 'Lansia', 'status lahan']
28.
29. X = df[kolom]
30. y = df.ket
31.
32. from sklearn.model_selection import train_test_split
33.
34. X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
    random_state=1)
35.
36. # Normalisasi menggunakan min-max : antara 0 - 1
37.
38. from sklearn.preprocessing import StandardScaler
39.
40. scaler = StandardScaler()
41. scaler.fit(X)
```

```
42.  
43. from sklearn.metrics import accuracy_score  
44. from sklearn.metrics import precision_score  
45. from sklearn.metrics import recall_score  
46. from sklearn.metrics import f1_score, classification_report  
47. from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay  
48. import matplotlib.pyplot as plt  
49.  
50. print('Performa\n')  
51. print('Akurasi :',accuracy_score(y_test, y_pred))  
52. print('Precision :',precision_score(y_test, y_pred, average='weighted'))  
53. print('Recall :',recall_score(y_test, y_pred, average='weighted'))  
54. print('F1-Score :',f1_score(y_test, y_pred, average='weighted'))  
55. print("\nConfusion Matrix :\n")  
56. cm = confusion_matrix(y_true = y_test, y_pred = y_pred, labels=rf.classes_)  
57. disp = ConfusionMatrixDisplay(confusion_matrix=cm,  
    display_labels=rf.classes_)  
58. disp.plot()  
59.  
60.     plt.show()
```

