PENERAPAN ALGORITMA *DECISION TREE* DALAM PEMBERIAN REKOMENDASI BANTUAN SISWA MISKIN (BSM) DI SEKOLAH DASAR SWASTA ARSYADIAH

Mariyani, Yulita Molliq Rangkuti, Arnah Ritonga
Ilmu Komputer, Universitas Negeri Medan
Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut Sei Tuan
Kabupaten Deli Serdang, Sumatera Utara 20221
mariyanimawardi@gmail.com

ABSTRAK

Pemerintah memiliki kewajiban untuk memberikan kesempatan seluasnya dan memberi kemudahan kepada masyarakat untuk mengikuti pendidikan sampai tamat SMA (Sekolah Menengah Atas). Bantuan Siswa Miskin (BSM) adalah Program Nasional yang bertujuan untuk membantu meringankan siswa miskin untuk bersekolah dengan bantuan akses pelayanan pendidikan yang layak. Pemberian BSM (Bantuan Siswa Miskin) disalurkan oleh Kementrian Pendidikan dan Kebudayaan melalui Direktorat Pembinaan Sekolah Dasar kepada setiap SD (Sekolah Dasar) diseluruh Indonesia. Pengumpulan data bantuan siswa miskin di SD Swasta Arsyadiah Medan masih menggunakan cara yang manual sehingga pemberian bantuan dilakukan dalam waktu yang lama (±3-4 minggu) Masalah lain yang ditemukan adalah adanya bantuan yang tidak tepat sasaran sehingga ada siswa yang kurang mampu namun tidak diberikan bantuan oleh pihak sekolah. Penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang memengaruhi rekomendasi bantuan siswa miskin dan mengevaluasi penerapan algoritma Decision Tree di Sekolah Dasar Swasta Arsyadiah. Data primer dikumpulkan melalui survei dan wawancara, sementara data sekunder berasal dari Sekolah Dasar Harapan Mulia. Algoritma Decision Tree diterapkan pada dataset siswa kelas 1-6 di Arsyadiah, dengan evaluasi menggunakan teknik 5 K-Fold Cross Validation. Model di Arsyadiah memiliki akurasi 63.8%, sementara di Harapan Mulia akurasinya 94.4%. Uji *one-tail t-test* menunjukkan perbedaan signifikan, mengindikasikan model lebih akurat pada data sekunder. Penelitian menyimpulkan bahwa faktor-faktor spesifik memengaruhi rekomendasi, dan algoritma ini efektif dalam memberikan rekomendasi bantuan meskipun akurasinya berbeda pada dataset yang digunakan.

Kata kunci : *Decision Tree, Sekolah Dasar Swasta Arsyadiah, Bantuan, Kemiskinan*

1. PENDAHULUAN

Program Bantuan Siswa Miskin (BSM) adalah program nasional yang bertujuan untuk membantu siswa dari keluarga miskin agar tetap bersekolah, mencegah putus sekolah, dan menarik siswa yang telah berhenti untuk kembali belajar. Program ini bertujuan meringankan biaya pendidikan dan memenuhi kebutuhan belajar, agar siswa dari keluarga kurang mampu dapat melanjutkan pendidikan hingga tamat. BSM tidak berdasarkan prestasi, tetapi mempertimbangkan kondisi ekonomi siswa. Program ini disalurkan oleh Kementerian Pendidikan melalui sekolah-sekolah, termasuk SD Swasta Arsyadiah Medan, yang memberikan bantuan uang tunai Rp. 450.000 setiap enam bulan [1].

Pengumpulan data atau berkas- berkas bantuan siswa miskin di SD Swasta Arsyadiah medan masih menggunakan cara yang manual dalam memilih satu persatu yang berhak menerima bantuan siswa miskin (BSM). Prosedur pengolahan data yang dilakukan meliputi kegiatan pengumpulan data, pengelompokan, pencocokan data dengan biodata siswa, perkiraan siswa penerima, dan menyusun laporan. Sehingga pemberian bantuan dilakukan cukup lama yaitu kurang lebih mencapai 3 sampai 4 minggu. Kemudian saat pemberian dana bantuan banyak siswa yang protes karena tidak tepat sasaran, ada siswa yang memang kurang mampu namun tidak diberikan bantuan oleh pihak sekolah Hal ini akan membuat pihak sekolah kewalahan dalam menyeleksi belum lagi yang penuh dengan subjektifitas dan kesalahan, sehingga dikhawatirkan pemilihan penerima (BSM) di sekolah tersebut tidak dibagikan secara adil dan tepat. Sehingga, penting untuk diteliti untuk mengetahui pola penentuan siswa penerima bantuan siswa miskin dapat dilakukan dengan cara memanfaatkan data siswa dan data penerima bantuan siswa miskin. Pencarian pola penentuan siswa penerima bantuan siswa miskin diharapkan bisa menganalisa faktor-faktor yang sangat menpengaruhi pada penerima bantuan siswa miskin dan bisa membantu pihak sekolah menentukan lebih berhak yang mendapatkan bantuan dana tersebut. Metode yang sesuai dengan permasalahan diatas adalah dengan menggunakan Metode decison tree. Decision tree diharapkan dapat diterapkan dalam kasus penerimaan bantuan siswa miskin berdasarkan keadaan ekonomi dan bisa mendapat hasil lebih akurat dan juga bisa mendapatkan siswa yang lebih tepat untuk mendapatkan bantuan siswa miskin [2].

2. TINJAUAN PUSTAKA

2.1. Decision Tree

Algoritma decision tree adalah metode pengambilan keputusan berbasis pohon keputusan

(tree) yang digunakan dalam machine learning dan data mining. Decision tree sendiri adalah sebuah struktur data berbentuk pohon yang memiliki akar (root), cabang (branch), dan daun (leaf), dimana setiap cabang dan daun memiliki keputusan atau label yang merepresentasikan hasil klasifikasi [3].

Langkah-langkah dalam membangun algoritma decision tree meliputi pemahaman masalah dan tujuan pemodelan, pemilihan data yang sesuai, serta persiapan data dengan memilih fitur relevan, menangani missing value, dan normalisasi jika diperlukan. Selanjutnya, fitur dipilih sebagai variabel independen, data dibagi pada node root berdasarkan nilai fitur yang dipilih, dan pembagian sub-tree dilakukan hingga tercapai kondisi pemutusan. Kelas sampel ditentukan dengan mengikuti jalur dari node root hingga leaf node. Model kemudian divalidasi untuk menilai akurasinya, diikuti dengan fine-tuning jika diperlukan, dan akhirnya model digunakan untuk memprediksi kelas data baru [4].

Dalam membangun decision tree untuk rekomendasi bantuan siswa miskin, langkah pertama adalah menentukan variabel target, seperti "menerima bantuan" atau "tidak menerima bantuan." Information gain kemudian dihitung untuk setiap variabel, seperti penghasilan orang tua, pekerjaan orang tua, jumlah tanggungan keluarga, surat keterangan tidak mampu, dan nilai rata-rata rapor, untuk memilih variabel dengan gain tertinggi sebagai node root. Misalnya, jika penghasilan orang tua memiliki gain tertinggi, itu akan dijadikan node root, dan data akan dibagi menjadi cabang "rendah" dan "tinggi." Proses ini berlanjut dengan menghitung gain variabel lain seperti jumlah anggota keluarga, tingkat kehadiran, dan tingkat kesehatan siswa pada setiap cabang hingga data homogen. Penghitungan gain dilakukan dengan mengukur entropy atau ketidakpastian sebelum dan sesudah pembagian data, yang bertujuan menghasilkan decision tree yang akurat dalam memprediksi layak atau tidaknya siswa menerima bantuan [5].

$$Entropy(S) = \sum_{i=1}^{n} -pi \log_2 pi \qquad (1)$$

Keterangan :
S : Himpunan Kasus
N : Jumlah partisi S
Pi : Proporsi S_i terhadap S

Pada Decision Tree digunakan untuk mengukur berapa banyak informasi yang diperoleh (information gain) atau berkurang (information loss) saat melakukan pemisahan node berdasarkan suatu atribut atau fitur tertentu. Information Gain biasanya digunakan untuk memilih atribut terbaik untuk membagi data dalam pembangunan pohon keputusan [6]. Rumus Information Gain adalah sebagai berikut :

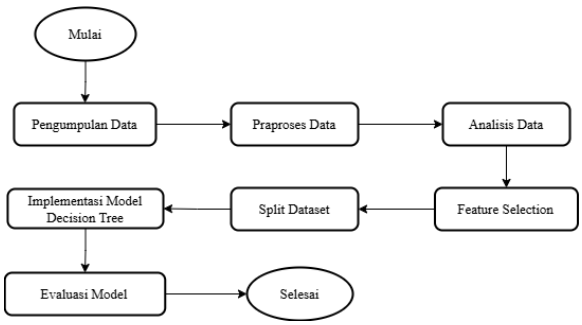
$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy(S_i) \quad (2)$$

2.2. Confusion Matrix

Confusion Matrix adalah alat yang digunakan untuk mengukur kinerja model klasifikasi dalam memprediksi kelas atau label yang berbeda, terdiri dari empat komponen utama: True Positives (TP), yaitu kasus di mana model benar memprediksi kelas positif; True Negatives (TN), ketika model benar memprediksi kelas negatif; False Positives (FP), yaitu kasus di mana model salah memprediksi kelas positif; dan False Negatives (FN), yaitu saat model salah memprediksi kelas negatif. Confusion Matrix membantu memahami akurasi model dalam memprediksi kelas dengan tepat dan mendeteksi kesalahan dalam klasifikasi [6].

K-Fold Cross Validation adalah teknik validasi yang membagi dataset menjadi K subset (fold) untuk menguji kemampuan generalisasi model dan menghindari overfitting. Pada setiap iterasi, satu subset digunakan sebagai data pengujian, sedangkan sisanya sebagai data pelatihan, dan ini dilakukan bergantian hingga semua subset digunakan sebagai validation set. Setelah setiap iterasi, kinerja model dievaluasi menggunakan metrik seperti akurasi, presisi, recall, atau F1-score, dan hasil dari semua iterasi dirata-rata untuk memberikan gambaran yang lebih akurat tentang kinerja model secara keseluruhan [7].

3. METODE PENELITIAN



Gambar 1. Flowchart Metode Penelitian

Tahap-tahap penelitian dapat dilihat pada Gambar 1. Penelitian dimulai dengan mengumpulkan data. Penelitian dilakukan di SD Swasta Arsyadiah, yang berlokasi di Jl. Sei Deli No. 139 Medan dan SD Harapan Mulia di Jl. Marelan 7, Pasar 1 Tengah, Terjun, Medan Marelan. Data yang telah dikumpulkan terdiri dari 1165 siswa untuk dataset primer (Tabel 1) dan 228 siswa untuk dataset sekunder (Tabel 2). Data primer mencakup seluruh siswa di SD Arsyadiah, sementara data sekunder diambil dari SD Harapan Mulia.

Tabel 1. Data Primer

Tahun	Jumlah Sampel
2020	291
2021	291
2022	291
2023	292
Jumlah	**1165**

Tabel 2. Data Sekunder

Tahun	Jumlah Sampel
2020	57
2021	57
2022	57
2023	57
Jumlah	**228**

Berikut di bawah ini adalah tabel pemilihan kriteria tentang penerapan algoritma *decision tree* penentuan rekomendasi bantuan siswa miskin di sekolah Arsyadiah Medan.

Tabel 3. Pemilihan Kriteria Siswa

No	Variabel	Jenis Data
1	Pekerjaan Orang Tua	Kategorik
2	Penghasilan Orang Tua	Numerik
3	Jumlah Tanggungan Keluarga	Numerik
4	Surat Keterangan Tidak Mampu	Kategorik
5	Hasil Nilai rata-rata rapot	Numerik

Tahap selanjutnya melakukan preprocessing data yang mencakup tahap pembersihan (cleaning), di mana data yang tidak valid, duplikat, atau tidak diperlukan dihapus. Selanjutnya, dibuat model decision tree untuk menentukan rekomendasi bantuan siswa miskin. Pembuatan model ini dilakukan dengan memanfaatkan algoritma decision tree, di mana data yang telah diproses akan diolah menjadi pohon keputusan. Setelah model decision tree selesai, dilakukan evaluasi untuk memastikan keefektifan model tersebut dalam memberikan rekomendasi bantuan. Evaluasi ini dilakukan dengan membandingkan hasil prediksi dari model dengan hasil aktual menggunakan confusion matrix [7].

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan dua jenis dataset, yaitu dataset primer dan sekunder. Dataset primer terdiri dari 1165 data yang dikumpulkan selama empat tahun, yaitu 2020 hingga 2023, sedangkan dataset sekunder mencakup 228 data dari periode yang sama. Semua data tersebut disusun dalam file Excel. Pengumpulan informasi dilakukan dari berbagai sumber yang relevan dengan tujuan penelitian. Setelah data terkumpul, dilakukan pre-processing menggunakan library pandas untuk menyiapkan data yang lebih terstruktur dan siap dianalisis. Tahap pre-processing mencakup pembersihan data, transformasi variabel, dan penghapusan outlier, dengan tujuan meningkatkan kualitas data sehingga dapat digunakan untuk analisis lebih lanjut.

Pada tahap pembersihan data, langkah awal yang dilakukan adalah menghilangkan nilai kosong (NaN) menggunakan fungsi drop.na() dari pandas. Langkah ini penting untuk menjaga konsistensi dan kualitas dataset. Gambar 1 menunjukkan visualisasi dataset sebelum proses pembersihan, sementara Gambar 2 menampilkan hasil setelah proses pembersihan, di mana baris atau kolom dengan nilai NaN telah dihapus. Proses ini memastikan dataset yang lebih konsisten dan dapat diandalkan untuk analisis lebih lanjut. Dengan data yang sudah bersih, hasil analisis diharapkan lebih valid, yang akan berkontribusi positif terhadap kesimpulan penelitian.

	No	Nama Siswa	Pekerjaan orang tua	Penghasilan Orang Tua	Jumlah Tanggungan Orang Tua	Nilai Rata - Rata	Surat Keterangan Miskin (Dapat dan Tidak Dapat)
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	1.0	Airin Maricka	T. Parkir	2200000.0	2.0	78.0	Dapat
2	2.0	Diandra Dinata	Wiraswasta	1800000.0	2.0	75.0	Dapat
3	3.0	Bintang Meydira	Jual Baju	1700000.0	3.0	70.0	Tidak Dapat
4	4.0	FazilaAzzahra	Wiraswasta	5200000.0	3.0	80.0	Tidak Dapat
...	...	...	...	...	...	...	...
1161	1161.0	jefri al diandra	Tukang becak	2400000.0	3.0	90.0	Tidak Dapat
1162	1162.0	rizki tri rafanda	Jualan Lontong	3000000.0	3.0	90.0	Tidak Dapat
1163	1163.0	hasan rifqi	Nelayan	2700000.0	1.0	92.0	Dapat
1164	1164.0	stella	Nelayan	3500000.0	3.0	85.0	Tidak Dapat
1165	1165.0	raihan dwiyansyah	Buruh cuci	2900000.0	2.0	85.0	Tidak Dapat

Gambar 2. Contoh Dataset Sebelum Dilakukan Cleaning

	No	Nama Siswa	Pekerjaan orang tua	Penghasilan Orang Tua	Jumlah Tanggungan Orang Tua	Nilai Rata - Rata	Surat Keterangan Miskin (Dapat dan Tidak Dapat)
1	1.0	Airin Maricka	T. Parkir	2200000.0	2.0	78.0	Dapat
2	2.0	Diandra Dinata	Wiraswasta	1800000.0	2.0	75.0	Dapat
3	3.0	Bintang Meydira	Jual Baju	1700000.0	3.0	70.0	Tidak Dapat
4	4.0	FazilaAzzahra	Wiraswasta	5200000.0	3.0	80.0	Tidak Dapat
5	5.0	M.Wahyu Saputra	Wiraswasta	2800000.0	2.0	82.0	Dapat
...	...	...	...	...	...	...	...
1161	1161.0	jefri al diandra	Tukang becak	2400000.0	3.0	90.0	Tidak Dapat
1162	1162.0	rizki tri rafanda	Jualan Lontong	3000000.0	3.0	90.0	Tidak Dapat
1163	1163.0	hasan rifqi	Nelayan	2700000.0	1.0	92.0	Dapat
1164	1164.0	stella	Nelayan	3500000.0	3.0	85.0	Tidak Dapat
1165	1165.0	raihan dwiyansyah	Buruh cuci	2900000.0	2.0	85.0	Tidak Dapat

1165 rows × 7 columns

Gambar 3. Contoh Dataset Setelah Dilakukan Celaning

Setelah tahap feature selection, langkah selanjutnya dalam pemrosesan data pada penelitian ini adalah split dataset. Pemisahan dataset menjadi data training dan data testing bertujuan untuk melatih model pada data yang sudah ada dan menguji performanya pada data yang belum pernah dilihat sebelumnya [8].

Tabel 4. Split dataset data primer

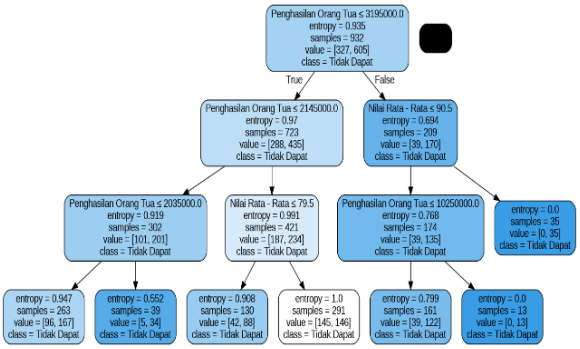
Split Dataset	
Training	935
Testing	230
Jumlah	1165

Tabel 5. Split dataset data sekunder

Split Dataset	
Training	182
Testing	46
Jumlah	228

Dataset yang telah dilakukan split dataset selanjutnya akan dilakukan metode K-Fold Cross Validation, K-Fold Cross Validation dilakukan pada dataset primer dan dataset sekunder. Pada penelitian ini digunakan nilai K-Fold Cross Validation sebesar 5. tujuan dilakukannya K-Fold Cross Validation pada penelitian ini adalah untuk mengukur kinerja algoritma decision tree secara lebih umum dan mencegah overfitting [9].

4.1. Decision Tree Dataset Primer



Gambar 1. Visualisasi Model Decision Tree Dataset Primer

4.2. Node Pertama (Root):

Jika Penghasilan Orang Tua (s) kurang dari atau sama dengan 3,195,000, maka keputusan pertama adalah "Dapat" dengan entropy sebesar 0.935. Total sampel yang mencapai node ini adalah 932, dengan pembagian kelas [327 Dapat, 605 Tidak Dapat].

Entropy (Penghasilan $\leq$ 3,195,000)

$= -\left(\frac{327}{932}\right)\log_2(\frac{327}{932}) - (\frac{605}{932})\log_2(\frac{605}{932})$

$= 0.935$

4.3. Cabang Pertama (True):

Jika Penghasilan Orang Tua (s) kurang dari atau sama dengan 2,145,000 maka keputusan tetap "Dapat" dengan entropy 0.97. Sampel yang mencapai node ini adalah sebanyak 723, dengan pembagian kelas [288 Dapat, 435 Tidak Dapat].

Entropy (Penghasilan $\leq$ 2,145,000)

$= -\left(\frac{288}{723}\right)\log_2(\frac{288}{723}) - (\frac{435}{723})\log_2(\frac{435}{723})$

$= 0.97$

4.4. Cabang Kedua (True):

Jika nilai rata-rata rapot lebih kecil atau sama dengan 79.5, maka keputusan adalah "Dapat" dengan entropy sebesar 0.991. Sampel yang mencapai node ini adalah sebanyak 421, dengan pembagian kelas [187 Dapat, 234 Tidak Dapat].

Entropy (Nilai Rata $-$ Rata $\leq$ 79.5)

$= -\left(\frac{187}{421}\right)\log_2(\frac{187}{421}) - (\frac{234}{421})\log_2(\frac{234}{421})$

$= 0.991$

4.5. Rule

Rule 1: Jika Penghasilan Orang Tua $\leq$ 2145000.0 dan Nilai Rata-Rata $\leq$ 90.5, maka kelasnya adalah "Tidak Dapat".

Rule 2: Jika Penghasilan Orang Tua $\leq$ 2145000.0 dan Nilai Rata-Rata > 90.5, maka lanjut ke aturan berikutnya.

Rule 3: Jika Penghasilan Orang Tua $\leq$ 2035000.0 dan Nilai Rata-Rata $\leq$ 79.5, maka kelasnya adalah "Tidak Dapat".

Rule 4: Jika Penghasilan Orang Tua $\leq$ 2035000.0 dan Nilai Rata-Rata > 79.5, maka lanjut ke aturan berikutnya.

Rule 5: Jika Penghasilan Orang Tua $\leq$ 10250000.0 dan Entropy = 0.768, maka kelasnya adalah "Tidak Dapat".

Rule 6: Jika Penghasilan Orang Tua $\leq$ 10250000.0 dan Entropy = 0.947, maka kelasnya adalah "Tidak Dapat".

Rule 7: Jika Penghasilan Orang Tua $\leq$ 10250000.0 dan Entropy = 0.552, maka kelasnya adalah "Tidak Dapat".

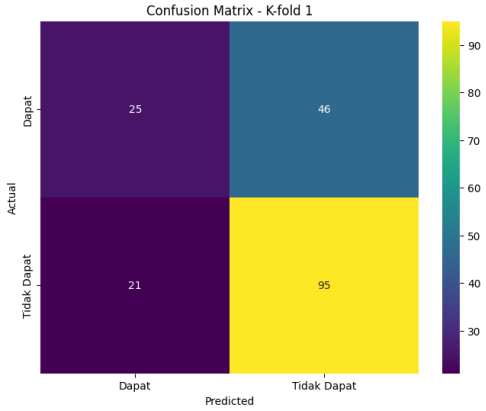
Rule 8: Jika Penghasilan Orang Tua $\leq$ 10250000.0 dan Entropy = 0.908, maka kelasnya adalah "Tidak Dapat".

Rule 9: Jika Penghasilan Orang Tua $\leq$ 10250000.0 dan Entropy = 0.0, maka kelasnya adalah "Tidak Dapat".

Rule 10: Jika Penghasilan Orang Tua $\leq$ 10250000.0 dan Entropy = 1.0, maka kelasnya adalah "Tidak Dapat".

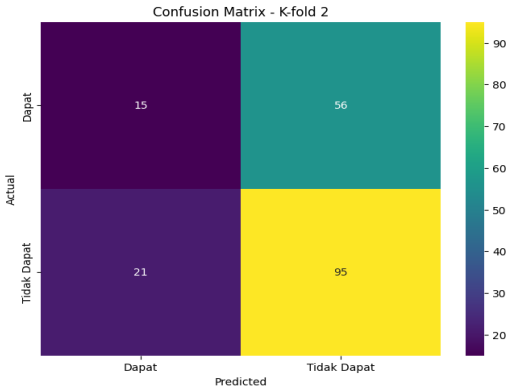
Rule 11: Jika Penghasilan Orang Tua $\leq$ 10250000.0 dan Entropy = 0.799, maka kelasnya adalah "Tidak Dapat".

Rule 12: Jika Penghasilan Orang Tua $\leq$ 10250000.0 dan Entropy = 0.0, maka kelasnya adalah "Tidak Dapat".



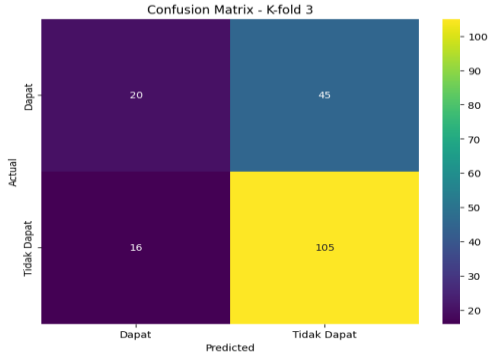
Gambar 2. Confusion Matrix K-Fold 1 Dataset Primer

Gambar 5 menunjukkan hasil confusion matrix dari model K-Fold 1 pada Dataset Primer, dimana terdapat 25 data *true positive*, 95 *true negative*, 46 *false negative* dan 21 *false positive*.



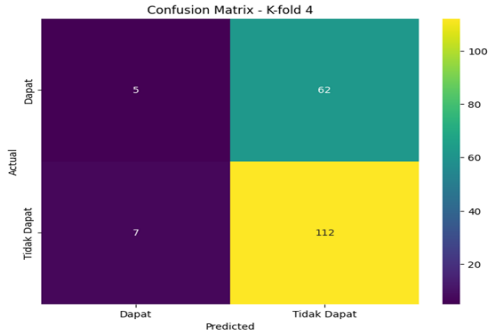
Gambar 3. Confusion Matrix K-Fold 2 Dataset Primer

Gambar 6 menunjukkan hasil confusion matrix dari model K-Fold 2 pada Dataset Primer, dimana terdapat 15 data *true positive*, 95 *true negative*, 56 *false negative* dan 21 *false positive*.



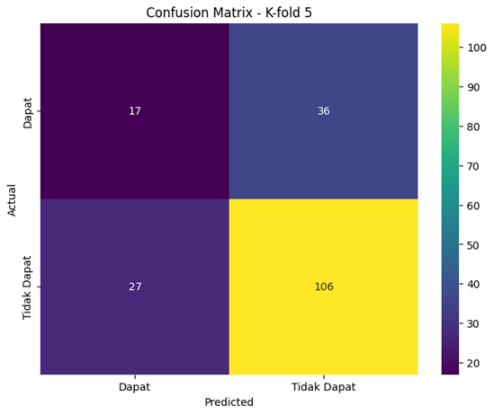
Gambar 4. Confusion Matrix K-Fold 3 Dataset Primer

Gambar 7 menunjukkan hasil confusion matrix dari model K-Fold 3 pada Dataset Primer, dimana terdapat 20 data *true positive*, 105 *true negative*, 16 *false negative* dan 45 *false positive*.



Gambar 5. Confusion Matrix K-Fold 4 Dataset Primer

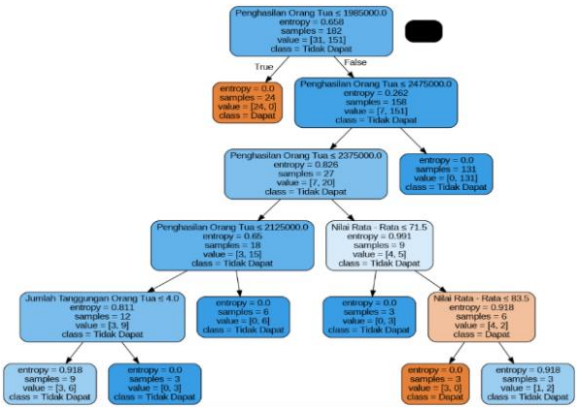
Gambar 8 menunjukkan hasil confusion matrix dari model K-Fold 4 pada Dataset Primer, dimana terdapat 5 data *true positive*, 112 *true negative*, 7 *false negative* dan 62 *false positive*.



Gambar 6. Confusion Matrix K-Fold 5 Dataset Primer

Gambar 9 menunjukkan hasil confusion matrix dari model K-Fold 5 pada Dataset Primer, dimana terdapat 17 data *true positive*, 106 *true negative*, 27 *false negative* dan 36 *false positive*.

4.6. Decision Tree Dataset Sekunder



Gambar 5. Visualisasi Model Decision Tree Dataset Sekunder

1. Node Pertama (Root):
   Jika Penghasilan Orang Tua (s) kurang dari atau sama dengan 1,985,000, maka keputusan pertama adalah "Tidak Dapat" dengan entropy sebesar 0.658. Total sampel yang mencapai node ini adalah 182, dengan pembagian kelas [31 Dapat, 151 Tidak Dapat].
   $Entropy\ (Penghasilan \leq 1{,}985.000)$
   $= -\left(\frac{31}{182}\right)\log_2\left(\frac{31}{182}\right) - \left(\frac{151}{182}\right)\log_2\left(\frac{151}{182}\right)$
   $= 0.658$

2. Cabang Pertama (True):
   Jika Penghasilan Orang Tua (s) kurang dari atau sama dengan 2,475,000 maka keputusan pertama adalah "Dapat" dengan entropy sebesar 0.0. Sampel yang mencapai node ini adalah 24, dengan pembagian kelas [24 Dapat, 0 Tidak Dapat].
   $Entropy\ (Penghasilan \leq 2{,}475{,}000)$
   $= -\left(\frac{24}{24}\right)\log_2\left(\frac{24}{24}\right) - \left(\frac{0}{24}\right)\log_2\left(\frac{0}{24}\right)$
   $= 0$

3. Cabang Kedua (True):
   Jika Penghasilan Orang Tua (s) kurang dari atau sama dengan 2,375,000 maka keputusan pertama adalah "Tidak Dapat" dengan entropy sebesar 0.826. Sampel yang mencapai node ini adalah 27, dengan pembagian kelas [7 Dapat, 20 Tidak Dapat].
   $Entropy\ (Penghasilan \leq 2{,}375{,}000)$
   $= -\left(\frac{7}{27}\right)\log_2(\frac{7}{27}) - (\frac{20}{27})\log_2(\frac{20}{27})$
   $= 0.866$

4. Cabang Ketiga (True):
   Jika Penghasilan Orang Tua (s) kurang dari atau sama dengan 2,125,000 maka keputusan pertama adalah "Tidak Dapat" dengan entropy sebesar 0.65. Sampel yang mencapai node ini adalah 18, dengan pembagian kelas [3 Dapat, 15 Tidak Dapat].
   $Entropy\ (Penghasilan \leq 2{,}125{,}000)$
   $= -\left(\frac{3}{18}\right)\log_2(\frac{3}{18}) - (\frac{15}{18})\log_2(\frac{15}{18})$
   $= 0.65$
   Jika Nilai Rata - Rata $\leq$ 71.5 maka keputusan adalah "Tidak Dapat" dengan entropy sebesar 0.991. Sampel yang mencapai node ini adalah sebanyak 9, dengan pembagian kelas [4 Dapat, 5 Tidak Dapat].

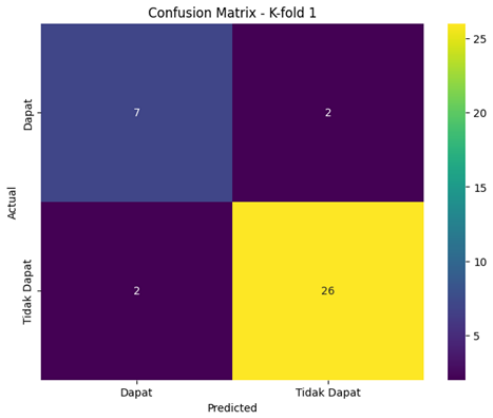
$Entropy\ (\text{Nilai Rata} - \text{Rata} \leq 71.5)$
$= -\left(\frac{4}{9}\right)\log_2(\frac{4}{9}) - (\frac{5}{9})\log_2(\frac{5}{9})$
$= 0.991$

Jika Jumlah Tanggungan Orang Tua $\leq$ 4.0 maka keputusan adalah "Tidak Dapat" dengan entropy sebesar 0.811. Sampel yang mencapai node ini adalah sebanyak 12, dengan pembagian kelas [3 Dapat, 9 Tidak Dapat].
$Entropy\ (\text{Jumlah Tanggungan Orang Tua} \leq 4.0)$
$= -\left(\frac{3}{12}\right)\log_2(\frac{3}{12}) - (\frac{9}{12})\log_2(\frac{9}{12})$
$= 0.811$

Jika Nilai Rata - Rata $\leq$ 83.5 maka keputusan adalah "Dapat" dengan entropy sebesar 0.918. Sampel yang mencapai node ini adalah sebanyak 9, dengan pembagian kelas [4 Dapat, 5 Tidak Dapat].
$Entropy\ (\text{Nilai Rata} - \text{Rata} \leq 83.5)$
$= -\left(\frac{4}{9}\right)\log_2(\frac{4}{9}) - (\frac{5}{9})\log_2(\frac{5}{9})$
$= 0.918$

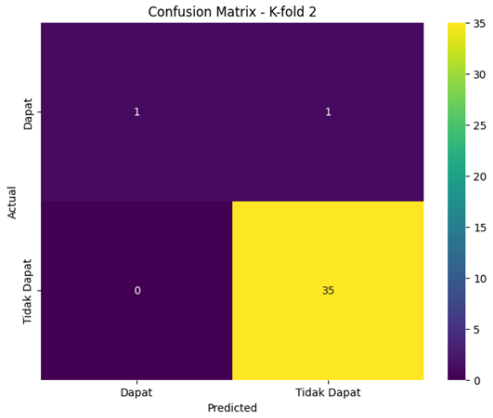
5. Rule :

Rule 1: Jika Penghasilan Orang Tua $\leq$ 1985000.0, maka kelasnya adalah "Tidak Dapat".
Rule 2: Jika Penghasilan Orang Tua $\leq$ 2475000.0, maka kelasnya adalah "Dapat".
Rule 3: Jika Penghasilan Orang Tua $\leq$ 2375000.0 dan Nilai Rata - Rata = 71.5, makakelasnya adalah "Tidak Dapat".
Rule 4: Jika Penghasilan Orang Tua $\leq$ 2125000.0 dan Jumlah Tanggungan Orang Tua = 4.0, maka kelasnya adalah "Tidak Dapat".
Rule 5: Jika Nilai Rata - Rata $\leq$ 71.5, maka kelasnya adalah "Tidak Dapat".
Rule 6: Jika Nilai Rata - Rata $\leq$ 83.5, maka kelasnya adalah "Dapat".



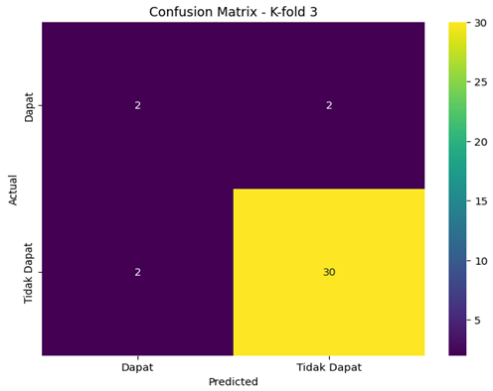
Gambar 7. Confusion Matrix K-Fold 1 Dataset Sekunder

Gambar 11 menunjukkan hasil confusion matrix dari model K-Fold 1 pada Dataset Sekunder, dimana terdapat 7 data *true positive*, 26 *true negative*, 2 *false negative* dan 2 *false positive*.



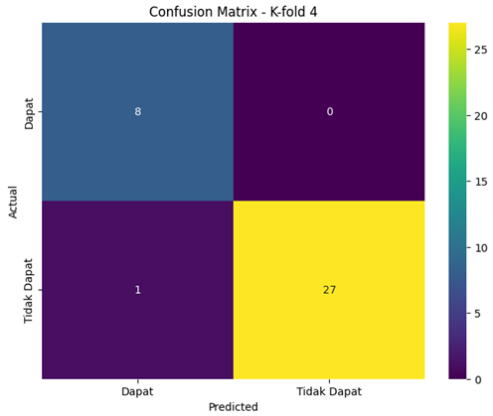
Gambar 8. Confusion Matrix K-Fold 2 Dataset Sekunder

Gambar 12 menunjukkan hasil confusion matrix dari model K-Fold 2 pada Dataset Sekunder, dimana terdapat 1 data *true positive*, 35 *true negative*, 1 *false negative* dan 0 *false positive*.



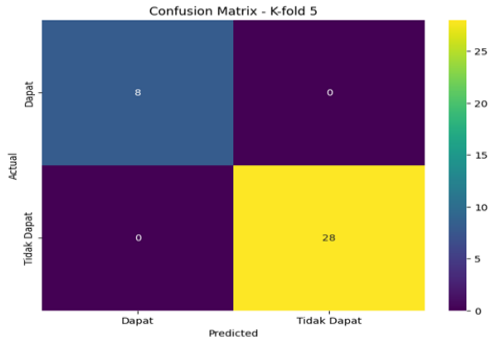
Gambar 9. Confusion Matrix K-Fold 3 Dataset Sekunder

Gambar 13 menunjukkan hasil confusion matrix dari model K-Fold 3 pada Dataset Sekunder, dimana terdapat 2 data *true positive*, 30 *true negative*, 2 *false negative* dan 2 *false positive*.



Gambar 10. Confusion Matrix K-Fold 4 Dataset Sekunder

Gambar 14 menunjukkan hasil confusion matrix dari model K-Fold 4 pada Dataset Sekunder, dimana terdapat 8 data *true positive*, 0 *true negative*, 1 *false negative* dan 27 *false positive*.



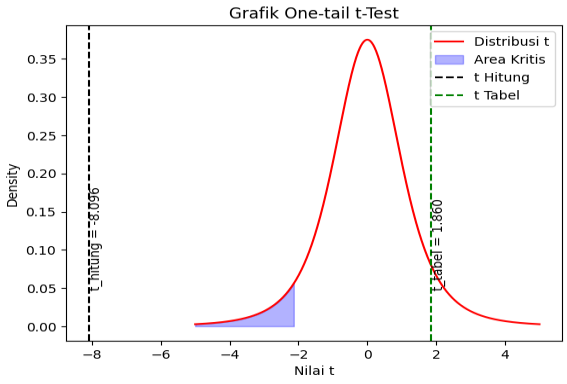
Gambar 11. Confusion Matrix K-Fold 5 Dataset Sekunder

Gambar 15 menunjukkan hasil confusion matrix dari model K-Fold 5 pada Dataset Primer, dimana terdapat 8 data *true positive*, 28 *true negative*, 0 *false negative* dan 0 *false positive*.

Tabel 6. Perbandingan Akurasi Dataset Primer Dan Sekunder

Akurasi	***Primer***	***Sekunder***
K-Fold 1	64%	89%
K-Fold 2	59%	97%
K-Fold 3	67%	89%
K-Fold 4	63%	97%
K-Fold 5	66%	100%
Rata-Rata	**63.8%**	**94.4%**

Tabel 6 memberikan gambaran hasil evaluasi kinerja suatu model atau metode klasifikasi melalui penerapan teknik validasi silang K-Fold dengan lima iterasi K-Fold yang berbeda. Dalam proses validasi silang K-Fold, data dibagi menjadi k subset dengan ukuran yang sama, di mana model dievaluasi sebanyak k kali. Setiap iterasi K-Fold melibatkan satu subset sebagai data pengujian sementara k-1 subset lainnya digunakan sebagai data pelatihan. Pada masing-masing iterasi K-Fold, dua metrik evaluasi yang digunakan adalah akurasi primer dan akurasi sekunder. Akurasi primer mengukur persentase prediksi yang benar secara keseluruhan, sementara akurasi sekunder mungkin merupakan pengukuran yang lebih spesifik atau terkait dengan jenis kesalahan tertentu yang lebih penting untuk diperhatikan. Variasi hasil dari masing-masing iterasi K-Fold menggambarkan keragaman dalam kinerja model, dengan rentang nilai akurasi primer dari 59% hingga 67% dan nilai akurasi sekunder dari 89% hingga 100% [10].



Gambar 12. Grafik Distribusi t uji signifikansi

Gambar 6. menjelaskan nilai kritis untuk tingkat signifikansi 0.05 dengan derajat kebebasan 8 adalah **1.860** dan nilai t yang dihitung adalah **-8.096**. Dari grafik dapat dilihat bahwa nilai t hitung tidak berada didalam kurva dan jauh dari nilai t tabel. Dalam hal ini dapat disimpulkan bahwa hipotesis nol (H0) ditolak, maka H1 diterima, yaitu terdapat perbedaan yang signifikan antara rata-tara akurasi dataset primer dan juga akurasi dataset sekunder dengan metode k-fold cross validation.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, didapatkan bahwa faktor yang harus dipertimbangkan dalam menentukan rekomendasi bantuan siswa miskin di SD Swasta Arsyadiah adalah "Penghasilan orang tua" dan "Nilai Rapot Siswa". Akurasi rata-rata model pada dataset primer adalah 63.8%, sedangkan pada dataset sekunder adalah 94.4%, menunjukkan bahwa akurasi dataset sekunder jauh lebih baik. Selain itu dapat disimpulkan juga bahwa model yang telah dibuat lebih mampu melakukan klasifikasi pada siswa yang "Tidak mendapatkan surat keterangan miskin" dibandingkan siswa yang mendapatkan "Surat keterangan miskin". Adapun saran untuk penelitian mendatang adalah menggunakan algoritma yang berbeda seperti Naive Bayes, Random Forest, atau Support Vector Machine, melakukan normalisasi pada dataset, menggunakan kriteria yang lebih banyak, menambah jumlah dataset, serta melakukan *deploy model* ke dalam bentuk *website* atau aplikasi.

DAFTAR PUSTAKA

[1] I. F. Prayoga, "Sistem Pendukung Penerimaan Beasiswa Sekolah menggunakan Metode K-Nearest Neighbor (KNN) Di SMA PGRI," *Teknologipintar.org*, vol. 3, no. 10, pp. 2023–2024, Mar. 2023.

[2] Y. Muamar and A. Muhajirin, "Penerapan Jaringan Saraf Tiruan Dengan Metode Backpropagation Untuk Memprediksi Tingkat Kelulusan Mahasiswa Perguruan Tinggi," *Digital Transformation Technology*, vol. 4, no. 1, pp. 214–224, 2024.

[3] A. R. Raharja, A. Pramudianto, Y. Muchsam, and others, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data 'Framingham' Untuk Menunjukkan Risiko Seseorang Terkena Penyakit Jantung dalam 10 Tahun Mendatang," *Technologia Journal*, vol. 1, no. 1, 2024.

[4] A. I. Rizmayanti, N. Hidayati, F. S. Nugraha, W. Gata, and S. N. Mandiri, "Penerapan Data Mining Untuk Memprediksi Kompetensi Siswa Menggunakan Metode Decission Tree (Studi Kasus Smk Multicomp Depok)," *Jurnal Swabumi*, vol. 9, no. 1, p. 2021, 2021.

[5] S. V. Natasya, R. M. Awangga, and others, "Penerapan Decision Tree (ID3) Untuk Profiling Mahasiswa dan Alumni," *SMATIKA JURNAL: STIKI Informatika Jurnal*, vol. 13, no. 02, pp. 250–259, 2023.

[6] M. Solehuddin, W. A. Syafei, and R. Gernowo, "Metode Decision Tree untuk Meningkatkan Kualitas Rencana Pelaksanaan Pembelajaran dengan Algoritma C4.5," *Jurnal Penelitian dan Pengembangan Pendidikan*, vol. 6, no. 3, pp. 510–519, 2022, doi: 10.23887/jppp.v6i3.52840.

[7] I. Markoulidakis, G. Kopsiaftis, I. Rallis, and I. Georgoulas, "Multi-Class Confusion Matrix Reduction method and its application on Net Promoter Score classification problem," *ACM International Conference Proceeding Series*, no. Cx, pp. 412–419, 2021, doi: 10.1145/3453892.3461323.

[8] B. Vrigazova, "The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems," *Business Systems Research*, vol. 12, no. 1, pp. 228–242, 2021, doi: 10.2478/bsrj-2021-0015.

[9] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *Jurnal Algoritma*, vol. 21, no. 1, pp. 239–248, 2024.

[10] S. Bates, T. Hastie, and R. Tibshirani, "Cross-validation: what does it estimate and how well does it do it?," *J Am Stat Assoc*, vol. 119, no. 546, pp. 1434–1445, 2024