

Decision Tree dalam Menganalisis Data Pengunjung Wisata Danau Poso untuk Pengambilan Keputusan

Fredryc Joshua Pa'o, Hendry*

Fakultas Teknologi Informasi, Program Studi Teknik Informatika, Universitas Kristen Satya Wacana, Salatiga, Indonesia

Email: ¹fredriycjosh@email.com, ²*hendry@uksw.edu

Submitted: 26/02/2021; Accepted: 09/04/2021; Published: 30/05/2021

Abstrak—Penelitian ini menggunakan sistem klasifikasi dalam mengelolah datanya. Dalam klasifikasi ada beberapa metode yang disediakan salah satunya adalah metode decision tree dengan algoritma C4.5 metode ini memiliki arti yaitu pohon keputusan dimana strukturnya sama dengan flowchart dimana setiap simpul menandakan suatu tes atribut, setiap cabang mempresentasikan hasil tes dan simpul daun mempresentasikan kelas atau distribusi kelas. Data yang dipakai adalah data pengunjung Wisata Danau Poso yang berada pada tahun 2009 sampai 2020, kemudian metode yang dipakai dalam penelitian ini dibagi menjadi beberapa tahap yaitu pengumpulan data, menganalisis data, cleaning data, transformasi data dan merancang decision tree dengan algoritma C4.5. Hasil yang dicapai dari penelitian ini adalah jumlah pengunjung lebih dari 28.984 memiliki keterangan “Banyak” yang didominasi oleh wisatawan lokal sedangkan nilai dengan keterangan “Kurang” ada pada wisatawan mancanegara. Hal ini menjadi salah satu poin penting dalam menetapkan strategi yang tepat untuk mengembangkan wisata di danau poso.

Kata Kunci: Klasifikasi; Decision Tree; Algoritma C4.5; Wisatawan Danau Poso.

Abstract—This study uses a classification system in managing its data. In classification there are several methods provided, one of which is the decision tree method with the C4.5 algorithm this method means a decision tree where the structure is the same as a flowchart where each node signifies an attribute test, each branch presents the test results and the leaf node represents the class or class distribution. The data used is the data of Lake Poso Tourism visitors from 2009 to 2020, then the method used in this study is divided into several stages, namely the data being studied, analyzing the data, transforming data and designing a decision tree with the C4.5 algorithm. The results achieved from this study are that the number of visitors more than 28,984 has a description of "Much" which is dominated by local tourists, while the value with the name "Less" is in foreign tourists. This is one of the important points in determining the right strategy for developing tourism in Lake Poso.

Keywords: Classification; Decision Tree; C4.5 Algorithm; Lake Poso Tourists

1. PENDAHULUAN

Pada zaman sekarang negara Indonesia tidak hanya dikenal dengan sumber kekayaan alam yang melimpah tetapi juga dikenal dengan tempat wisatanya yang menjadikan Indonesia salah satu negara yang paling banyak dikunjungi oleh banyak orang, keindahan dan keunikan yang menjadi salah satu ciri khas dari tempat wisata di negara kita membuat tempat wisata di Indonesia sudah banyak dikenal oleh banyak orang baik dari dalam negeri maupun dari luar negeri selain itu ada sejarah khusus yang menjadi daya tarik tersendiri bagi wisatawan [1]. Salah satu tempat wisata yang banyak dikunjungi adalah tempat wisata di danau poso, tempat ini berada di pulau Sulawesi tepatnya di daerah provinsi Sulawesi Tengah kabupaten Poso, ada banyak tempat wisata yang dapat dikunjungi di danau poso selain keindahannya banyak orang yang ingin mengetahui sejarah terbentuknya tempat wisata danau poso. Pemerintah setempat menjadikan wisata danau poso untuk menjadi salah satu penghasilan daerah dan lapangan pekerjaan. Maka dari itu pemerintah daerah kabupaten poso berusaha membangun tempat wisata danau poso menjadi lebih baik lagi [2].

Pendataan yang dilakukan oleh pemerintah terbilang cukup efektif, ada 2 kantor yang menjadi pusat pendataan pengunjung wisata danau poso, namun dari data yang dikumpulkan pemerintah setempat tidak menggunakannya dengan efektif padahal dari data yang dikumpulkan pemerintah setempat dapat dijadikan salah satu poin penting dalam menetapkan strategi pengembangan wisata danau poso seperti dalam penelitian terdahulu yaitu “Strategi Pengembangan Wisata Danau Poso Kecamatan Pamona Pusalemba Kabupaten Poso”, namun penelitian ini masi belum menggunakan data yang ada secara yang efektif untuk menghasilkan sebuah keputusan guna mengatur strategi pengembangan Wisata Danau Poso. [3]. Dengan adanya penelitian ini maka akan membantu dalam mengolah data yang sudah dikumpulkan dan mengelolanya dengan efektif kemudian akan memberikan hasil sehingga pemerintah setempat akan punya banyak strategi yang dapat diterapkan untuk memajukan wisata danau poso [4].

Dalam penelitian ini menggunakan metode Decision Tree yang merupakan salah satu teknik klasifikasi yang berguna untuk mempelajari dan memprediksi pola dari data yang menggambarkan relasi dari variabel yang satu dengan variabel yang lainnya. Metode ini akan memakai data yang sudah di kumpulkan lalu mengolahnya sehingga tercipta suatu kesimpulan yang akan dipakai untuk membantu menetapkan strategi atau rencana yang akan dipakai [5].

Berdasarkan latar belakang yang ada penelitian ini akan membantu mengolah data yang sudah dikumpulkan dari dua kantor pariwisata danau poso, data yang berhasil dikumpulkan adalah pengunjung data wisata danau poso 12 tahun terakhir. Keakuratannya memang tidak sempurna tapi melalui penelitian ini dapat membantu untuk pengambilan suatu keputusan yang cukup efektif [6].

2. METODE PENELITIAN

2.1 Sistem Pendukung Penelitian

Data Mining adalah ilmu yang mempelajari tentang pemecahan suatu masalah dengan menganalisis data yang tersedia dalam database dan menghasilkan sebuah keputusan, selain itu Data Mining juga untuk menemukan informasi baru dengan teknik mencari pola atau aturan tertentu dari sejumlah data dalam jumlah yang besar. Data Mining memiliki salah satu teknik yang disebut dengan klasifikasi [7]. Klasifikasi sendiri adalah suatu proses yang menemukan properti-properti yang mirip dalam sebuah kumpulan objek dalam sebuah data dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang berbeda yang bertujuan untuk menemukan model dari training set yang membedakan atribut ke dalam kategori atau kelas yang sesuai model tersebut akan digunakan untuk mengklasifikasi atribut yang kelasnya belum diketahui sebelumnya [8].

Decision Tree dijuluki sebagai pohon keputusan memiliki algoritma yang disebut dengan algoritma C4.5 yang merupakan salah satu algoritma yang digunakan untuk membentuk decision tree berdasarkan training data [9]. Algoritman ini menggunakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Konsep dari decision tree yaitu merubah data menjadi decision tree dan aturan-aturan keputusan atau disebut dengan decision rules dan merupakan pengembangan dari Algoritma ID3 [10]. Ada beberapa tahap yang digunakan dalam penerapan decision tree dengan Algoritma C4.5, berikut tahapannya [11]:

1. Mempersiapkan data training, biasanya data yang diambil adalah data historis atau data masa lalu yang sudah dikelompokkan ke dalam kategori tertentu.
2. Menghitung akar dari pohon. Akar akan diambil dari atribut yang dipilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dahulu nilai entropy. Rumus untuk menghitung nilai entropy :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i \log_2 p_i \quad (1)$$

Keterangan:

- S = Himpunan kasus
- n = Jumlah partisi S
- p_i = Proporsi S_i terhadap S

Kemudian hitung nilai gain menggunakan rumus:

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

Keterangan:

- S = Himpunan kasus
- A = Fitur
- n = Jumlah partisi atribut A
- $|S_i|$ = Proporsi S_i terhadap S
- $|S|$ = Jumlah kasus dalam S

3. Ulangi langkah ke 2 dan langkah ke 3 sehingga semua record terpartisi
4. Proses partisi decision tree akan berhenti saat :
 Semua record dalam simpul N mendapat kelas yang sama.
 1. Tidak ada atribut didalam record yang dipartisi lagi.
 2. Tidak ada record didalam cabang yang kosong.

2.2 Pengumpulan Data

Penelitian ini menggunakan data yang dikumpulkan dari 2 kantor di kota berbeda yaitu pada kantor pariwisata kota Poso dan kota Tentena, diberikan langsung oleh supervisor bagian pengelolaan data dan data yang diberikan adalah data dari tahun 2009 sampai dengan tahun 2020.

2.3 Menganalisis Data

Setelah data terkumpul maka diperlukan proses menganalisa data terlebih dahulu, data yang digunakan adalah data pengunjung wisata danau poso. Dari hasil analisa tersebut telah ditentukan atribut yang akan dipakai yaitu Tahun, Wisatawan Mancanegara, Wisatawan Lokal, Jumlah seluruh Wisatawan. Beberapa dari empat atribut diatas akan dijadikan input untuk menghasilkan atribut target dimana atribut target tersebut akan menjadi class output untuk menentukan banyak atau kurangnya jumlah pengunjung berikut ini adalah tabel samples data yang akan digunakan.

Tabel 1. Data Pengunjung Wisata Danau Poso

Tahun	Wisatawan Mancanegara	Wisatawan Loka	Jumlah
2009	1.874	7.249	9.123
2010	2.228	9.073	11.301
2011	2.887	9.062	11.949

Tahun	Wisatawan Mancanegara	Wisatawan Loka	Jumlah
2012	3.367	12.408	15.775
2013	3.427	25.521	28.948
2014	3.352	9.035	12.387
2015	4.329	42.426	46.755
2016	3.061	79.098	82.159
2017	1.450	86.540	87.990
2018	1.017	15.506	16.523
2019	3.555	110.263	113.818
2020	416	23.836	24.252

3. HASIL DAN PEMBAHASAN

3.1 Cleaning Data

Penelitian ini menggunakan data pengunjung wisata danau poso. Dalam proses cleaning, data yang akan digunakan sebagai atribut di model decision tree yang akan dirancang menggunakan algoritma C4.5 adalah WISATAWAN MANCANEGARA, WISATAWAN LOKAL dan Jumlah. Sedangkan untuk atribut kelas atau output dapat ditentukan berdasarkan tahun pengunjung yang didata.

3.2 Mentransformasi Data

Setelah dilakukannya proses cleaning, selanjutnya adalah melakukan proses mentransformasi data. Pertama-tama adalah membagi data perempat tahun dan mengklasifikasinya yang terletak pada atribut tahun kemudian mencari rata-rata dari data dalam atribut jumlah sebagai patokan kurang atau banyaknya pengunjung.

Tabel 2. Hasil Klasifikasi Data

Tahun	Wisatawan Mancanegara	Wisatawan Lokal	Jumlah
2009	1.874	7.249	9.123
2010	2.228	9.073	11.301
2011	2.887	9.062	11.949
2012	3.367	12.408	15.775
2013	3.427	25.521	28.948
2014	3.352	9.035	12.387
2015	4.329	42.426	46.755
2016	3.061	79.098	82.159
2017	1.450	86.540	87.990
2018	1.017	15.506	16.523
2019	3.555	110.263	113.818
2020	416	23.836	24.252
Rata-rata			38.415

Berdasarkan tabel 2, sudah diketahui rata-rata dan hasil klasifikasinya maksud dari pengelompokan diatas adalah untuk menentukan apakah pengunjung pada atribut WISATAWAN MANCANEGARA dan WISATAWAN LOKAL masuk pada kategori kurang atau banyak berdasarkan rata-rata yang telah didapat pada atribut jumlah. Selanjutnya dari hasil diatas maka akan dilakukan analisa menggunakan algoritma C4.5 untuk menentukan kurang atau banyaknya pengunjung sesuai atribut target.

Tabel 3. Atribut Target

Atribut Target	keterangan
Kurang	≤ 38.415
Banyak	> 38.415

Setelah proses tabel 2 selesai maka diketahui ada 12 sample yang sudah di klasifikasi berdasarkan rata-rata yang sudah ditentukan diatribut jumlah dan setelah ditransformasi dapat menghasilkan keputusan banyak dan kurang Seperti tabel 4.

Tabel 4. Hasil Transformasi Data

Tahun	Wisatawan Mancanegara	Wisatawan Lokal	Jumlah	Ket
2009	1.874	7.249	9.123	Kurang
2010	2.228	9.073	11.301	Kurang
2011	2.887	9.062	11.949	Kurang
2012	3.367	12.408	15.775	Kurang

Tahun	Wisatawan Mancanegara	Wisatawan Lokal	Jumlah	Ket
2013	3.427	25.521	28.948	Kurang
2014	3.352	9.035	12.387	Kurang
2015	4.329	42.426	46.755	Banyak
2016	3.061	79.098	82.159	Banyak
2017	1.450	86.540	87.990	Banyak
2018	1.017	15.506	16.523	Kurang
2019	3.555	110.263	113.818	Banyak
2020	416	23.836	24.252	Kurang
Rata-rata				38.415

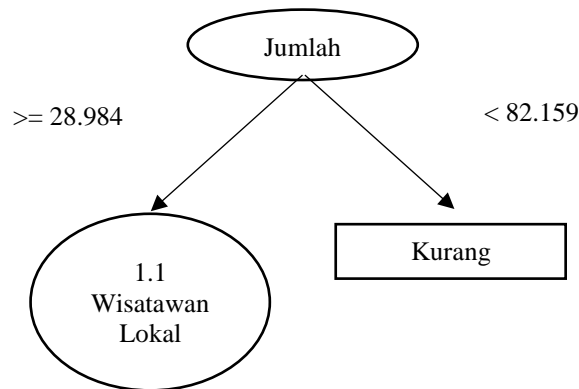
3.3 Merancang Decision Tree menggunakan Algoritma C4.5

Berdasarkan tabel 4, dilakukan proses perhitungan entropy (1) dan gain (1) guna menentukan akar (root) dari pohon keputusan dalam menentukan banyak kurangnya jumlah pengunjung. Berikut hasil perhitungan entropy dan gain node 1.

Tabel 5. Hasil Perhitungan Node 1

Node	Atribut	Value	Jumlah Kasus (S)	Kurang (S)	Banyak (S)	Entropy	Gain	Split Info	Gain Ratio
1	Total		12	8	4	0,918295			
						83405	0,612	0,9182	0,666
							1972	958340	6666
							227	5	6667
	Tahun	2009-2012	4	4	0	0			
		2013-2016	4	2	2	1			
		2017-2020	4	2	2	1			
							0,251	1,5849	0,158
							6291	625007	7603
							6738	2	2792
Wisatwan Lokal	>= 42.426	4	0	4	0				
	< 110.263	11	8	3	0,845350				
						1,693	0,6433	2,631	
						2008	908088	6833	
						5928	2	1917	
Wisatwan Mancanegara	>= 417	12	12	0	0				
	< 1.874	3	3	0	0				
						0,918		1,836	
						2958	0,5	5916	
						3405		681	
Jumlah	>= 28.984	8	4	4	1				
	< 82.159	7	6	1	0,591672				
						27858			
						0,596	0,0636	9,378	
						7716	294204	8630	
						2156	1	7536	

Dari hasil pada tabel 5, dapat diketahui bahwa atribut dengan gain tertinggi adalah Jumlah yaitu 9,37886307536 dengan begitu Jumlah akan dijadikan node akar. Dari hasil perhitungan nilai entropy dan gain pada node 1, dapat digambarkan decision tree sementara seperti berikut.



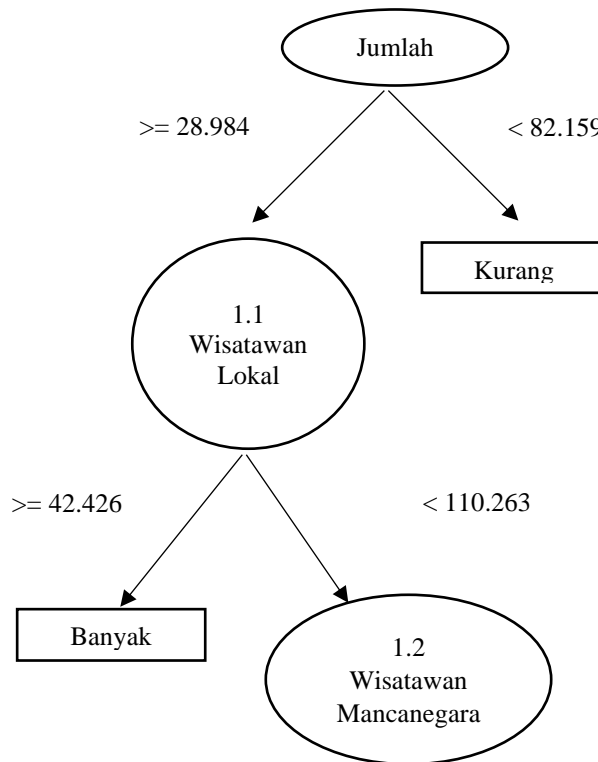
Gambar 1. Decision Tree Node 1

Hasil pada gambar 1 menunjukkan jumlah sebagai node akar, didalam jumlah ada 6 nilai atribut yang tetapi berdasarkan sampel data yang sudah dianalisis maka ada 2 klasifikasi yang digunakan pada atribut jumlah yaitu wisatawan ≥ 28.984 dan wisatawan < 82.159 . untuk wisatawan < 82.159 sudah mengklasifikasikan nilai menjadi satu keputusan yaitu “Kurang”, sedangkan wisatawan ≥ 28.984 masih belum mengklasifikasikan nilai menjadi satu keputusan, karena itu masih perlu dilakukan proses perhitungan entropy dan gain lebih lanjut hasilnya terlihat pada tabel 6 berikut.

Tabel 6. Hasil Perhitungan Node 1.1

Node	Atribut	Value	Jumlah Kasus (S)	Kurang (S)	Banyak (S)	Entropy	Gain	Split Info	Gain Ratio
1.1	Total		8	4	4				
	Tahun	2013-2016	4	2	2	1	1	1	1
		2017-2020	4	2	2	1			
								0,668	1,4957
							1	5644	421235
								432	5
	Wisatwan Lokal	≥ 42.42	4	0	4	0			
		< 110.263	7	4	3	0,985228			
								1,362	0,668
							0746	5644	2,0373
							1903	432	123831
	Wisatwan Mancanegara	≥ 2.228	8	8	0	0			
		< 3.352	3	3	0	0			
								0,530	1,8845
							1	6390	201403
								6223	

Pada hasil tabel 6 kembali diketahui bahwa atribut dengan gain tertinggi adalah pada wisatawan lokal dengan nilai 2,0373123831. Dengan demikian dapat menjadi node cabang dari jumlah yang bernilai ≥ 28.984 . Dari hasil perhitungan nilai entropy dan gain pada node 1.1, berikut gambaran decision tree sementara terlihat pada gambar 2 berikut.



Gambar 2. Decision Tree Node 1.1

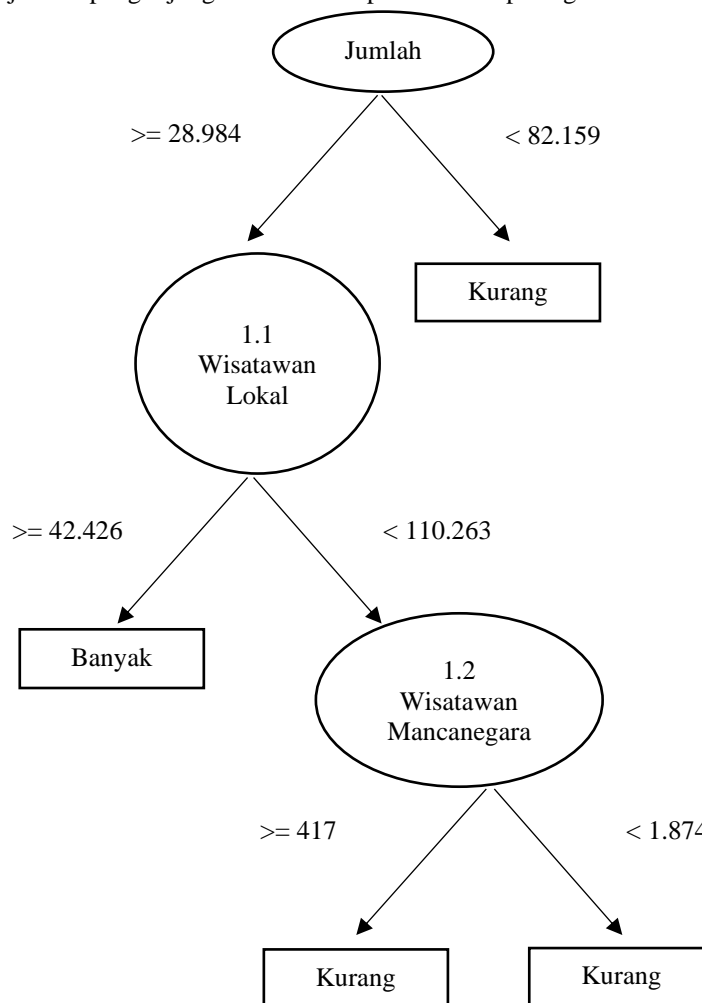
Atribut wisatawan memiliki 2 nilai atribut yang dipakai. Pada nilai atribut ≥ 42.426 sudah diklasifikasi menjadi satu keputusan yaitu “Banyak”, sedangkan nilai atribut “Wisatawan Mancanegara” belum diklasifikasi menjadi satu keputusan sehingga perlu dilakukan proses perhitungan kembali yaitu pada entropy dan gain dinode 1.2, berikut hasil perhitungannya.

Tabel 7. Hasil Node 1.2

Node	Atribut	Value	Jumlah Kasus (S)	Kurang (S)	Banyak(S)	Entropy	Gain	Split Info	Gain Ratio
1.2	Total		12	8	4	0,918295 83405	0,918 2958 3405	0,918 2958 3405	1
	Tahun	2009-2012	4	4	0	0			
		2013-2016	4	2	2	1			
		2017-2020	4	2	2	1			
							0,251 6291 6738	1,584 9625 0072	0,158 7603 2792
	Wisatawan Mancanegara	≥ 417	12	12	0	0			
		< 1.874	3	3	0	0			
							0,918 2958 3405	0,5	1,836 5916 681
		≥ 2.228	8	8	0	0			

Node	Atribut	Value	Jumlah Kasus (S)	Kurang (S)	Banyak(S)	Entropy	Gain	Split Info	Gain Ratio
		< 3.352	7	7	0	0	0,918	0,843	1,088
							2958	5794	5706
							3405	2137	9149
		>= 3.367	4	4	0	0			
		< 4.329	11	11	0	0			
							0,918	0,643	1,427
							2958	3908	2753
							3405	0882	3788

Dari hasil pada tabel 7 diketahui bahwa atribut dengan gain tertinggi ada pada atribut wisatawan “Wisatawan Mancanegara” yaitu pada nilai ≥ 417 dan < 1.874 . Dengan demikian atribut “Wisatawan Mancanegara” memiliki 2 nilai atribut yang sudah diklasifikasi dan menghasilkan satu keputusan yaitu nilai ≥ 417 menjadi “Kurang” dan nilai dari < 1.817 juga menjadi “kurang”. Karena kedua nilai dari atribut “Wisatawan Mancanegara” tersebut sudah mengklasifikasikan kasus menjadi 1. Maka digambarkan decision tree akhir dalam menentukan jumlah pengunjung wisata danau poso terlihat pada gambar 3 berikut.



Gambar 3. Decision Tree Node 1.2

Berdasarkan pengujian yang telah dilakukan, dapat disimpulkan bahwa hasil analisis penentuan pengunjung wisata danau poso dengan menggunakan algoritma C4.5 adalah sebagai berikut:

- a. Jumlah wisatawan dengan pengunjung kurang dari “82.159” maka keterangannya “Kurang” dan Jumlah lebih dari “28.984” maka maka keterangannya “Banyak” dan didominasi oleh atribut “Wisatawan Lokal”.

- b. Jumlah wisatawan lokal dengan pengunjung lebih dari atau sama dengan “42.426” maka keterangannya “Banyak” sedangkan pengunjung kurang dari “110.263” maka keterangannya “Kurang” dan berada pada atribut “Wisatawan Mancanegara”.
- c. Jumlah wisatawan Mancegara lebih dari sama dengan “>=417” dan kurang dari “1.874” keterangannya adalah “Kurang”.

4. KESIMPULAN

Dari perancangan Decision Tree dalam Menganalisis Data Pengunjung Wisata Danau Poso untuk Pengambilan Keputusan dapat disimpulkan bahwa pengunjung wisata danau poso dengan jumlah pengunjung lebih dari 28.984 memiliki keterangan banyak dan didominasi oleh wisatawan lokal sedangkan wisatawan mancanegara dengan jumlah pengunjung lebih dari 417 dan kurang dari 1.874 masih memiliki keterangan kurang. Hal ini menjadi perhatian pemerintah daerah setempat dalam membangun wisata danau poso sebagaimana diketahui wisatawan mancanegara masih belum banyak yang mengetahui letak wisata danau poso padahal potensi wisata danau poso sangat banyak jika diketahui oleh wisatawan mancanegara selain dapat menjadi tempat penghasilan dana daerah nama wisata danau poso juga akan banyak dikenal oleh orang asing dari luar Indonesia.

REFERENCES

- [1] V. Talimba *et al.*, “ISSN 2442-3262 KAJIAN DANAU POSO SEBAGAI DAERAH TUJUAN WISATA BERBASIS Abstrak Jurnal Perencanaan Wilayah dan Kota Jurnal Perencanaan Wilayah dan Kota,” vol. 7, no. 1, hal. 51–61, 2020.
- [2] D. Kebudayaan, D. Kabupaten, S. R. Primadany, dan P. Daerah, “ANALISIS STRATEGI PENGEMBANGAN PARIWISATA DAERAH,” vol. 1, no. 4, hal. 135–143.
- [3] A. Balingki, “Strategi pengembangan wisata di pesisir danau poso kecamatan pamona puselemba kabupaten poso,” *Strateg. Pengemb. WISATA DI PESISIR DANAU POSO Kec. PAMONA PUSELEMBE KABUPATEN POSO*, hal. 181–193, 2015.
- [4] I. G. A. K. Giantari, “Kata Kunci :Strategi, Pengembangan, Objek Wisata, Air Panas, Marobo ABSTRACT,” vol. 11, hal. 773–796, 2015.
- [5] I. Sutoyo, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik,” *J. Pilar Nusa Mandiri*, vol. 14, no. 2, hal. 217, 2018, doi: 10.33480/pilar.v14i2.926.
- [6] P. Kasih, “Pemodelan Data Mining Decision Tree Dengan Classification Error Untuk Seleksi Calon Anggota Tim Paduan Suara,” *Innov. Res. Informatics*, vol. 2, hal. 63–69, 2019.
- [7] Z. Azmi dan M. Dahria, “Decision Tree Berbasis Algoritma Untuk Pengambilan Keputusan,” *Saintikom*, vol. 12, hal. 157–164, 2013.
- [8] Rismayanti, “Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus : Fakultas Teknik dan Komputer Universitas Harapan Medan),” *J. Sist. Inf.*, vol. 02, no. 01, hal. 16–24, 2018.
- [9] Andie, “Penerapan Decision Tree Untuk Menganalisis Kemungkinan Pengunduran Diri Calon Mahasiswa Baru,” *Technologia*, vol. 7, no. 1, hal. 8–14, 2016.
- [10] Rismayanti, “Implementasi Algoritma C4.5 Untuk Menentukan Penerima Beasiswa Di Stt Harapan Medan,” *Media Infotama*, vol. 12, no. 2, hal. 116–120, 2016.
- [11] A. Andriani, “Sistem Prediksi Penyakit Diabetes Berbasis Decision Tree,” *Bianglala Inform.*, vol. 1, no. 1, hal. 1–10, 2013.