

ANALISA PAKET INDIHOME BERDASARKAN PENDAPATAN PERKAPITA SUMATERA MENGGUNAKAN RAPID MINER DENGAN ALGORITMA DECISION TREE

Corry Agustic¹

Universitas Pembangunan Panca Budi Medan

Email: corryagustic@gmail.com¹

Abstract: Indihome is one of the fixed broadband services that faces challenges in the form of differences in purchasing power of people in various regions, the packages offered are found to be incompatible with the income of the population. This study aims to recommend service packages that are appropriate to the income of the population based on per capita income as measured by the Gross Regional Domestic Product (GRDP) indicator per capita and average household expenditure per month using the Decion Tree method implemented in the Rapid Miner application. Data sourced from per capita income of 13 locations in Sumatera and a list of service packages. The results of the Decion Tree model study were able to classify into 3 parts, namely basic (20Mbps), Medium (50-75Mbps) and Premium (100-300Mbps) showing an accuracy of 92.73% in adjusting service packages and the average monthly expenditure variable being the most determining factor. These findings can be the basis for strategies in adjusting packages to the economy of a region.

Keywords: Decion Tree, Rapid Miner, Indihome, Per Capita Income, Sumatera

Abstrak: Indihome adalah salah satu layanan fixbradband yang menghadapi tantangan berupa perbedaan daya beli masyarakat diberbagai wilayah, paket yang ditawarkan ditemukan tidak sesuai dengan penghasilan penduduk. Penelitian ini bertujuan untuk merekomendasikan paket layanan yang sesuai dengan penghasilan penduduk berdasarkan pendapatan per kapita yang diukur dengan indikator Produk Domestik Regional Bruto (PDRB) per kapita dan rata-rata pengeluaran rumah tangga per bulan dengan menggunakan metode Decion Tree yang diimplementasikan ke dalam aplikasi Rapid Miner. Data bersumber dari Pendapatatn per kapita 13 lokasi di Sumatera dan daftar paket layanan. Hasil penelitian model Decion Tree mampu mengklasifikasikan menjadi 3 bagian yaitu basic (20Mbps), Medium (50-75Mbps) dan Premium (100-300Mbps) menunjukkan akurasi sebesar 92,73% dalam menyesuaikan paket layanan dan variabel rata-rata pengeluaran per bulan menjadi faktor yang paling menentukan. Temuan ini dapat menjadi dasar strategi dalam menyesuaikan paket terhadap ekonomi suatu wilayah.

Kata kunci: Decion Tree, Rapid Miner, Indihome, Pendapatan Per Kapita, Sumatera

PENDAHULUAN

Transformasi digital menempatkan internet diposisi paling penting yang menghubungkan banyak data menjadi kebutuhan dasar saat ini yang membawa kemudahan dalam melakukan aktifitas dibidang Pendidikan, Bisnis, Pemerintahan, Sosial. Indihome adalah salah satu merek layanan FTTH (Fiber To The Home) yang jaringan fiber optik nya

telah tergelar dan menjadi tulang punggung penyedia Broadband di Indonesia. Dengan demikian paket layanan yang ditawarkan tidak merata yang dipengaruhi oleh daya beli tingkat sosial ekonomi masyarakat.

Pulau Sumatera sebagai dataran ketiga terbesar di Indonesia memiliki banyak Provinsi dengan populasi masyarakat dan tingkat ekonomi yang heterogen. Dengan kondisi ini penawaran

paket layanan yang seragam akan tidak sesuai dengan pendapatan masyarakat. Paket kecepatan tinggi akan sulit disesuaikan dengan masyarakat yang berpenghasilan rendah dan masyarakat yang berpenghasilan tinggi akan dengan mudah memilih paket yang tersedia.

Dalam hal ini bagaimana menentukan paket layanan Indihome yang sesuai dengan pendapatan masyarakat disuatu wilayah dengan objektif dan terukur, dengan demikian diperlukan pendekatan berbasis data yang mampu menyesuaikan pendapatan per kapita menjadi rekomendasi paket layanan yang tepat.

Data mining mempunyai Teknik untuk mengklasifikasi salah satunya dengan algoritma Decision Tree (pohon keputusan) yang dapat mengolah data numerik dan kategorikal untuk pengambilan keputusan bisnis yang membutuhkan justifikasi tertentu. Rapid miner merupakan aplikasi berbasis operator yang dapat memproses data, pemodelan dan evaluasi. Kombinasi Rapid Miner dengan Decision Tree memungkinkan pembangunan model rekomendasi, evaluasi, adaptasi.

Tujuan penelitian ini adalah membuat model hubungan pendapatan per kapita dengan kesesuaian layanan paket, membangun dan evaluasi Decision Tree menggunakan Rapidminer, mengidentifikasi variabel sosial ekonomi yang paling menentukan dalam kesesuaian paket. Kontribusi penelitian terletak pada penyediaan klasifikasi yang dapat menjadikan strategi pemasaran dan

penetapan paket yang sesuai di wilayah Sumatera.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan kerangka kerja CRISP-DM (Cross Industri Standart Proses for Data Mining) terdiri dari pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi dan penerapan yang dijalankan di Rapid miner.

Sumber Data

Data yang digunakan adalah daftar paket yg terdiri dari 6 tingkatan kecepatan yang memiliki harga perbulan disajikan di Gambar.1, Data selanjutnya adalah data pendapatan per kapita yang mencakup 13 wilayah di Sumatera yang terdiri dari Provinsi, Kota, kabupaten dengan atribut PDRB per kapita rata-rata tahunan serta pengeluaran rumah tangga perbulan disajikan di Tabel.2



Gambar 1. Paket dan Harga layanan.

Tabel 1. Data Pendapatan Penduduk Per Kapita Wilayah Sumatera

Wilayah	Level	PDRB Per Kapita Tahunan (Ribu Rp)	Rata Pengeluaran Per Bulan (Rp)	Kelas_Ekonomi (Target/Label)
Kepulauan Riau	Provinsi	145.53	1950000	Tinggi
Riau	Provinsi	151.259	1650000	Tinggi
Jambi	Provinsi	76.224	1300000	Menengah
Sumatera Selatan	Provinsi	68.239	1250000	Menengah
Sumatera Utara	Provinsi	62.5	1200000	Menengah
Sumatera Barat	Provinsi	50.264	1220000	Menengah
Lampung	Provinsi	44.982	1100000	Rendah
Bengkulu	Provinsi	43.757	1050000	Rendah
Kota Medan	Kota	141.412	1850000	Tinggi
Kota Padang	Kota	92.21	1450000	Menengah-Atas
Deli Serdang	Kabupaten	78.077	1300000	Menengah
Kabupaten Agam	Kabupaten	52.226	1414557	Menengah
Nias Selatan	Kabupaten	24.242	850000	Rendah

Kesuaian Paket

Paket dibagi menjadi 3 bagian yaitu Basic (20Mbps), Medium (50-75Mbps) dan Premium (100-300Mbps). Label target ditentukan oleh prinsip keterjangkauan, suatu paket terjangkau apabila tidak melebihi 25% dari rata-rata pengeluaran rumah tangga setiap bulannya. Dari batasan tersebut tiap wilayah diarahkan ke paket tertinggi yang masih terjangkau.

Tabel 2. Pemetaan Kesesuaian Paket

Bagian	Paket Terjangkau	25% Pengeluaran
Basic	20 Mbps	< Rp262.500
Medium	50–75 Mbps	Rp262.500 – Rp327.450
Premium	100–300 Mbps	> Rp371.850

Augmentasi dan Persiapan Data

Dilakukan augmentasi data terhadap 13 wilayah Sumatera untuk memperoleh sample yang memadai untuk pengujian. Setiap wilayah menjadi acuan yang menggambarkan kecamatan/kluster rumah tangga dengan nilai PDRB per kapita dan pengeluaran secara acak mengikuti distribusi normal menghasilkan deviasi 10-20% proses ini menghasilkan 182 record.

Pemodelan Decision Tree pada Rapid Miner

Pembagian dataset ini menggunakan split data dengan rasio 70% (127 record) data latih dan 30% (55 record) data uji secara sampling. Model dibangun dengan operator Decision Tree yang menggunakan gain_ratio/gini index, kedalaman maksimum 4 dan minimum jumlah sampel 6 per daun mencegah overfitting. Alur operator rapid miner secara umum yaitu: Read excel – Set role (label) – Split data – Decision Tree – Apply Model – Performance (Clasification).

Tabel 3. Parameter Decision Tree

Parameter	Nilai
Kriteria pemisahan	Gini index/Gain ratio

Kedalaman maksimum (maximal depth)	4
Minimal sampel per daun (min leaf size)	6
Skema validasi	Split validation (holdout)
Rasio data latih : uji	70 : 30 (stratified)
Atribut prediktor	PDRB per kapita; Pengeluaran per bulan
Atribut target (label)	Bagian kesesuaian paket (3 Bagian)

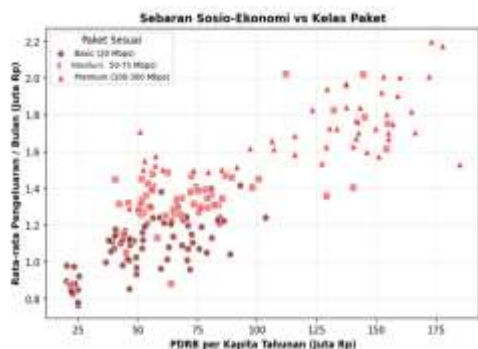
Metode Evaluasi

Kinerja model dievaluasi menggunakan confusion matrix untuk menampilkan benar dan salah untuk tiap bagian sehingga untuk memungkinkan pola analisis kesalahan klasifikasi. Akurasi mengukur proporsi prediksi benar terhadap seluruh data uji, sedangkan precision, recall, dan F1-score mengukur ketepatan model pada masing-masing bagian.

HASIL DAN PEMBAHASAN**Karakteristik dan Distribusi Bagian**

Hasil pelabelan menunjukkan distribusi paket relatif seimbang yaitu 74 Basic, 59 Medium, 49 Premium. Distribusi ini menggambarkan komposisi ekonomi Sumatera didominasi oleh pendapatan menengah kebawah dengan sebagian wilayah ekonomi berpenghasilan tinggi seperti Medan, Riau dan kepulauan Riau.

**Gambar 2. Distribusi Paket Layanan**



Gambar 3. Sebaran Sosial Ekonomi (PDRB vs Pengeluaran) terhadap Bagian Paket

Gambar 3 memperlihatkan bahwa pemisahan bagian cukup jelas berdasarkan penghasilan per bulan. Segmen basic pengeluaran <1.25Juta, Medium 1.25 – 1.49Juta dan Premium >1.49Juta. Pola ini menjadi variable awal bahwa pengeluaran memiliki pengaruh yang tinggi dengan paket.

Model Pohon Keputusan

Model Decision Tree menempatkan variabel pengeluaran per bulan sebagai akar. Aturan utama yang dihasilkan yaitu jika pengeluaran $\leq$ Rp1.249.000 maka paket direkomendasikan Basic, jika pengeluaran berada antara Rp1.249.000–Rp1.493.000 maka paket direkomendasikan Medium dan jika pengeluaran >Rp1.493.000 maka paket direkomendasikan Premium, dengan PDRB per kapita berperan sebagai pemisah sekunder pada beberapa cabang.



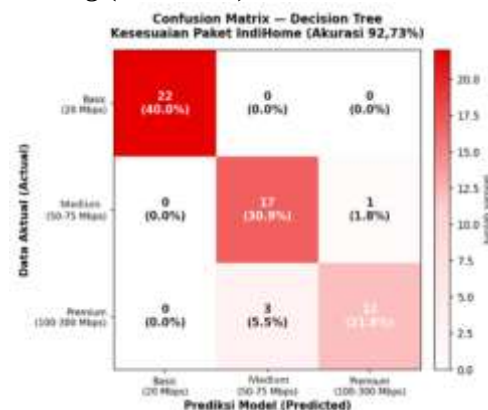
Gambar 4. Decision Tree hasil pemodelan

Paket yang direkomendasikan yaitu berdasarkan besarnya pengeluaran

bulanan masyarakat serta dapat langsung menjadi kebijakan paket disetiap wilayah.

Evaluasi model dan grafik matrix.

Evaluasi pada 55 data yang diuji menghasilkan akurasi keseluruhan sebesar 92,73%. Confusion matrix pada Gambar 5 memperlihatkan bahwa seluruh 22 sampel bagian Basic terklasifikasi secara sempurna. Kesalahan klasifikasi hanya terjadi di perbatasan bagian menengah atas, satu sampel Medium diprediksi sebagai Premium, dan tiga sampel Premium diprediksi sebagai Standar. Tidak ada kesalahan klasifikasi lintas dua tingkat, yang menunjukkan model konsisten dan hanya keliru pada kasus ambang (borderline).



Gambar 5. Confusion Matrix) model Decision Tree

Tabel 4. Confusion Matrix (Data Uji, n=55)

Prediksi	Premium	Medium	Basic	Total
Basic	0	0	22	22
Medium	1	17	0	18
Premium	12	3	0	15
Total	13	20	22	55

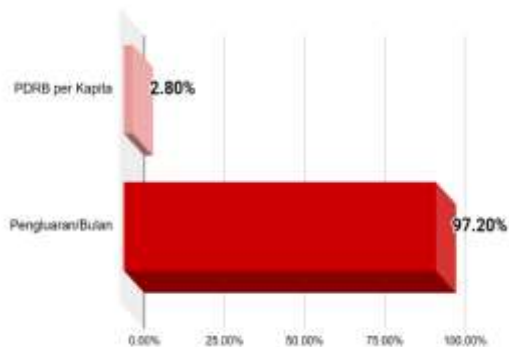
Rincian metrik per Bagian disajikan pada Tabel 5. Bagian Basic memperoleh precision, recall, dan F1-score sempurna (1,000). bagian Medium memiliki recall tinggi (0,944) namun precision sedikit lebih rendah (0,850) akibat menyerap sebagian sampel Premium. Sebaliknya, bagian Premium memiliki precision tinggi (0,923) tetapi recall lebih rendah (0,800). Rata-rata makro F1-score mencapai

0,917, mengonfirmasi kinerja model yang kuat dan seimbang.

Tabel 5. Metrix Kinerja Per Bagian

Bagian	Precision	Recall	F1-Score
Basic (20 Mbps)	1,000	1,000	1,000
Mdium (50–75 Mbps)	850	944	895
Premium (100–300 Mbps)	923	800	857
Rata-rata	924	915	917

Tingkat kepentingan variabel



Gambar 6. Feature Importance

Analisis feature importance pada Gambar 6 menunjukkan rata-rata pengeluaran per bulan merupakan variabel paling menentukan dengan kontribusi 97,2%, jauh melampaui PDRB per kapita (2,8%). Secara ekonomi pengeluaran rumah tangga merupakan cerminan daya beli yang mempengaruhi keputusan berlangganan.

SIMPULAN

Penelitian ini membangun model klasifikasi kesesuaian paket layanan indihome berdasarkan pendapatan per kapita penduduk Sumatera menggunakan algoritma Decision Tree pada Rapid Miner dengan mengklasifikasikan wilayah kedalam 3 bagian paket yaitu Basic, Medium dan Premium dengan akurasi 97,2% dengan variabel rata-rata pengeluaran per bulan terbukti menjadi faktor paling menentukan dengan 97,2%.

Model yang dihasilkan bersifat interpretatif, mudah diaudit, dapat langsung diterjemahkan menjadi kebijakan paket per wilayah. Dengan menyelaraskan penawaran paket terhadap kondisi sosial ekonomi maka penyedia layanan berpotensi meningkatkan keterjangkauan dan retensi pelanggan.

DAFTAR PUSTAKA

- Aditra Pradnyana, S.Kom., M.Kom, G., & Agustini, S.Si., M.Si., Dr. K. (2022). Konsep Dasar Data Mining. *Perpustakaan Digital Universitas Terbuka*, 498. <https://pustaka.ut.ac.id/lib/msim4403-data-mining/>
- Aggarwal, C. C. (2015). Data Mining: The Textbook. In *Data Mining: the Textbook*. <https://doi.org/10.1007/978-3-319-14142-8>
- Good, P. (2012). Lecture 11 Decision Trees. *A-Z of Error-Free Research*, 181–192.
- Mining, D. (2021). *Data Mining Model Overfitting Introduction to Data Mining , 2 nd Edition by Example Data Set Two class problem :*
- Rudin, C., Russell, C., & Berghen, V. (2012). *Decision Trees MIT*. 1–17. http://ocw.mit.edu/courses/sloan-school-of-management/15-097-prediction-machine-learning-and-statistics-spring-2012/lecture-notes/MIT15_097S12_lec08.pdf
- Tan, Steinbach, K. (2020). *Data Mining : Data Lecture Notes for Chapter 2 Introduction to Data Mining , 2 nd Edition by Attributes and Objects Types of Data Data Quality Similarity and Distance Data Preprocessing What is Data ? A More Complete View of Data Data may have parts Attri.*
- Tan, Steinbach, Karpatne, K. (2020). *Data Mining Classification : Basic Concepts and Techniques Lecture Notes for Chapter 3 Introduction to Data Mining , 2 nd Edition by Classification : Definition Given a collection of records (training set) – Each record is by characterized by a tuple (. 1–59.*