PREDIKSI INDEKS PEMBANGUNAN MANUSIA DENGAN MENGGUNAKAN *DECISION TREE C4.5* DI KABUPATEN KAMPAR

Ridwan Miftahul Putra, Elvira Asril, Taslim
Teknik Informatika, Universitas Lancang Kuning
Jl. Yos Sudarso No.KM.8, Rumbai, Pekanbaru, Riau
ridwanmiftahul12@gmail.com, elvira@unilak.ac.id, taslim@unilak.ac.id

ABSTRAK

Dengan perkembangan teknologi dalam bidang komputer yang pesat dapat mempermudah perolehan informasi. Pengolahan data yang disimpan tidak hanya bisa disimpan saja, tetapi data IPM bisa digunakan untuk kemudian hari. Dengan data mining, kita dapat melakukan pengklasifikasi, memprediksi, memperkirakan dan mendapatkan informasi yang lain yang bermanfaat dari kumpulan data dalam jumlah besar. Dalam IPM mempunyai beberapa variable yaitu angka harapan hidup, rata-rata lama sekolah, harapan lama sekolah, pengeluaran perkapita yang disesuaikan dan pengaturan range IPM. Memprediksi IPM dalam data mining dapat dilakukan dengan menggunakan metode *Decision Tree* dan algoritma C4.5. Dengan menggunakan algoritma C4.5 akan mendapatkan suatu hari berupa nilai entropy dan gain yang mudah dipahami dan mudah dimengerti.

1. PENDAHULUAN

Indeks Pembangunan Manusia (IPM) adalah mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup yang di bangun melalui pendekatan tiga dimensi dasar : dimensi panjang dan sehat yang diukur dari data angka harapan hidup, pengetahuan yang diukur dari data gabungan indikator angka huruf dan rata-rata lama sekolah, dan kehidupan yang layak yang diukur dari data indikator kemampuan daya beli masyarakat terhadap sejumlah kebutuhan pokok yang dilihat dari pengeluaran riil perKapita (Pratama & Al-Shaikh, 2012).

IPM ditujukan sebagai alat perencanaan dan evaluasi kebijakan pemerintah, seperti alokasi dan untuk daerah, sedangkan indikator IPM menggambarkan keberhasilan target pembangunan pemerintah. Sehingga dapat dikatakan, IPM menjadi data yang dapat digunakan dalam membuat kebijakan oleh pemerintah. Namun validitas data masih menjadi persoalan

klasik yang menghambat dalam membuat kebijakan, sejumlah instansi pemerintah pun memiliki data yang berbeda sehingga menciptakan kebingungan antar kementerian dalam membuat kebijakan dan pihak BPS mengumpulkan data masih secara manual, pihak BPS melakukan survei di tiap daerah untuk mendapatkan data dengan melakukan survei memerlukan waktu yang sangat lama dan membuat tenaga kerja yang banyak lebih dari 50 orang. Dari hasil wawancara saya dengan Bapak Ferwantono selaku bagian Tata Usaha di BPS Kampar, saya medapatkan permasalahan yang ada di Kantor BPS Kampar yang saya tuangkan di latar belakang.

Oleh sebab itu metode yang penulis pilih untuk memberikan solusi pada permasalahan penelitian ini yaitu dengan metode *decision tree C4.5*. Metode metode *decion tree* adalah pohon keputusan yang memperlihatkan faktor-faktor kemungkinan atau probabilitas yang akan mempengaruhi alternatif-

alternatif keputusan tersebut, disertai dengan estimasi hasil akhir yang akan didapat bila kita mengambil alternatif keputusan tersebut (Pramadhani & Setiadi, 2014). Variabel predictor yaitu angka harapan hidup, harapan lama sekolah, indek pembangunan manusia, pengeluaran perkapital dan rata-rata lama sekolah.

2. HASIL DAN PEMBAHASAN

a. Decision Tree

Decision Tree pengolahan data yang dapat diproses dan menghasilkan suatu keputusan yang saat berguna untuk mengeksplorasi serta menemukan hubungan tersembunyi antara sejumlah variable input dengan variable target. Dari suatu keputusan maka dapat memprediksi masa yang akan datang. Dengan memprediksi masa yang akan ada memudahkan mendapat informasi di tempat penelitian yang dilakukan. Seperti jumlah penduduk dalam setahun dan jumlah IPM.

b. Algoritma C4.5

Salah satu algoritma yang dapat digunakan untuk membuat pohon keputusan (*decission tree*) adalah algoritma C4.5.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut:

1. Pilih atribut sebagai akar
2. Buat cabang untuk tiap-tiap nilai
3. Bagi kasus dalam cabang
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama

Contoh penerapan Algoritma C4.5 :

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada.

Untuk menghitung gain digunakan persamaan 1.

$$\begin{aligned} &= Entropy S \\ &= \frac{|S_i|}{|S|} * Entropy S_i \end{aligned} \quad Gain$$

Keterangan :

S = himpunan kasus

A = atribut

N = Jumlah partisi

|S_i| = jumlah kasus pada partisi ke-i

|S| = jumlah kasus dalam S

Sementara itu, perhitungan nilai *entropy* dapat dilihat pada persamaan 2

$$Entropy S = \frac{n}{i=1} - p * \log_2 p_i$$

Keterangan :

S = himpunan kasus

A = atribut

n = jumkah partisi

Pi = proporsi dari Si terhadap S

Untuk lebih jelasnya langkah-langkah dalam pembuatan pohon keputusan, berikut diberikan contoh kasus yang dapat diklasifikasikan menggunakan algoritma C4.5.

Pada tahap ini permasalahan yang terjadi diatas akan dibahas dan diselesaikan berdasarkan data IPM pada table dibawah ini :

Tabel 2.1 IPM

Tahun	Angka Harapan Hidup	Rata-rata Lama Sekolah	Harapan Lama Sekolah	Pengeluaran Perkapita Yang disesuaikan	IPM (BPS)
2010	69,58	8,08	11,65	10,433	68.62
2011	69,65	8,56	11,93	10,513	69.64
2012	69,72	8,59	12,22	10,593	70.08
2013	69,77	8,62	12,51	10,624	70.46
2014	69,80	8,62	12,72	10,651	70.72

Perhitungan mendapatkan nilai IPM tahun 2010 :

Tabel 2.2 Nilai Maksimum dan Minimum Komponen IPM

Indikator Komponen IPM	Nilai Maksimum	Nilai Minimum
Angka Harapan Hidup	85	25
Harapan Lama Sekolah	100	0
Rata-rata Lama Sekolah	15	0
Pengeluaran Perkapita	732,72	300

Indeks angka harapan hidup :

$$(69,58 - 25) / (85 - 25) = 0,78210$$

Harapan lama sekolah : $(11,65 - 0)$

$$/ (100 - 0) = 0,1165$$

Indeks rata-rata lama sekolah :

$$(8,08 - 0) / (15 - 0) = 0,53866$$

Indeks pendidikan : $2/3$

$$(0,1165) + 1/3 (0,53866) = 0,27636$$

Indeks Konsumsi per kapita :

$$(10,433 - 300) / (732,72 - 300) = -0,66867$$

Akhirnya angka IPM dapat dihitung menggunakan persamaan :

$$\text{IPM} : 1/3 (\text{Indeks angka harapan hidup} + \text{Indeks pendidikan} + \text{Indeks konsumsi per kapita})$$

kapita)

$$: 1/3 (0,78210 + 0,27636 + -0,66867) =$$

$$0,11812$$

$$\text{IPM dalam ratusan} : 0,11812 * 100 =$$

$$11,8116$$

Tabel keterangan menentukan IPM level :

Tabel 2.3 IPM Level

Nilai Terendah	0,11035
Nilai Tertinggi	0,11812
Hasil	0,00259
High	0,11553
Medium	0,11294
Low	0,11035

Perhitungan mendapatkan IPM Level

$$\text{Hasil} = (\text{Nilai tertinggi} - \text{Nilai terendah})/3$$

$$= (0,11812 - 0,11035)/3$$

$$= 0,00259$$

$$\text{High} = \text{Nilai tertinggi} - \text{Hasil}$$

$$= 0,11812 - 0,00259$$

$$= 0,11553$$

$$\text{Medium} = \text{High} - \text{Hasil}$$

$$= 0,11553 - 0,00259$$

$$= 0,11294$$

$$\text{Low} = \text{Medium} - \text{Hasil}$$

$$= 0,11294 - 0,00259$$

$$= 0,11035$$

Tabel 2.4 Kategori

Nilai Terendah	0,11035
Nilai Tertinggi	0,11812
Hasil	0,003885
Good	0,114235
Bad	0,11035

Perhitungan mendapatkan kategori

$$\text{Hasil} = (\text{Nilai tertinggi} - \text{Nilai terendah})/2$$

$$\begin{aligned}
&= (0,11812 - 0,11035)/3 \\
&= 0,003885 \\
\text{Good} &= \text{Nilai tertinggi} - \text{Hasil} \\
&= 0,11812 - 0,003885 \\
&= 0,114235 \\
\text{Bad} &= \text{Good} - \text{Hasil} \\
&= 0,114235 - 0,003885 \\
&= 0,11035
\end{aligned}$$

Tabel 2.5 Hasil Analisa Berdasarkan Table Di Atas

Tahun	IPM	IPM Level	Kategori
2010	0.11035	Low	Bad
2011	0.11498	Medium	Good
2012	0.11629	High	Good
2013	0.11746	High	Good
2014	0.11812	High	Good

Tabel 2.6 Jumlah IPM Level dan Kategori

Good	4
Bad	1
Total	5
High	3
Medium	1
Low	1
High Good	3
Total High	3
High Bad	0
Total Bad	1
Medium Good	1
Total Medium	1
Medium Bad	0
Total Bad	1
Low Good	0
Total Low	1
Low Bad	1
Total bad	1

Penyelesaian :

$$\begin{aligned}
\text{Entropy} &= \sum_{i=1}^n -p_i \log_2 p_i \\
&= \frac{-\text{Good}}{\text{Total}} * \log_2 \left(\frac{\text{Good}}{\text{Total}} \right) + \\
&\quad \frac{-\text{Bad}}{\text{Total}} * \log_2 \left(\frac{\text{Bad}}{\text{Total}} \right)
\end{aligned}$$

$$\begin{aligned}
&= \frac{-4}{5} * \log_2 \left(\frac{4}{5} \right) + \frac{-1}{5} * \\
&\quad \log_2 \left(\frac{1}{5} \right) \\
&= -0,8 * \log_2(0,8) + \\
&\quad -0,2 * \log_2(0,2) \\
&= -0,8 * -0,321928095 + \\
&\quad -0,2 * -2,321928095 \\
&= 0,257542476 + \\
&\quad 0,464385619 \\
&= 0,72193
\end{aligned}$$

Nilai Gain High

$$\begin{aligned}
&= \frac{\text{High}}{\text{Total}} * \left(\frac{-\text{High Good}}{\text{Total High}} * \log_2 \frac{\text{High Good}}{\text{TotalH High}} \right) + \\
&\quad \frac{-\text{High Bad}}{\text{Total Bad}} * \log_2 \left(\frac{\text{High Bad}}{\text{Total Bad}} \right) \\
&= \frac{3}{5} * \left(\frac{-3}{3} * \log_2 * \frac{-3}{3} \right) + \frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \\
&= \frac{3}{5} * (-1 * 0) \\
&= 0,6 * 0 \\
&= 0
\end{aligned}$$

Nilai Gain Medium

$$\begin{aligned}
&= \frac{\text{Medium}}{\text{Total}} * \left(\frac{-\text{Medium Good}}{\text{Total Medium}} * \right. \\
&\quad \log_2 \frac{\text{Medium Good}}{\text{Total Medium}} \left. \right) + \frac{-\text{Medium Bad}}{\text{Total Bad}} * \\
&\quad \log_2 \left(\frac{\text{Medium Bad}}{\text{Total Bad}} \right) \\
&= \frac{1}{5} * \left(\frac{-1}{1} * \log_2 * \frac{1}{1} \right) + \frac{0}{1} * \log_2 \left(\frac{0}{1} \right) \\
&= \frac{1}{5} * (-1 * 0) \\
&= 0,2 * 0 \\
&= 0
\end{aligned}$$

Nilai Gain Low

$$\begin{aligned}
&= \frac{\text{Low}}{\text{Total}} * \left(\frac{-\text{Low Good}}{\text{Total Low}} * \log_2 \frac{\text{Low Good}}{\text{Total Low}} \right) + \\
&\quad \frac{-\text{Low Bad}}{\text{Total Bad}} * \log_2 \left(\frac{\text{Low Bad}}{\text{Total Bad}} \right) \\
&= \frac{1}{5} * \left(\frac{0}{1} * \log_2 * \frac{0}{1} \right) + \frac{-1}{1} * \log_2 \left(\frac{1}{1} \right)
\end{aligned}$$

$$= \frac{1}{5} * (0 * \log_2 0) + (-1 * \log_2 1)$$

$$= \frac{1}{5} * (0 + 0)$$

$$= 0$$

Nilai Gain Total / Hasil Prediksi

$$= \text{Entropy Total} - (\text{Nilai Gain} + \text{Nilai Medium} + \text{Nilai Low})$$

$$= 0,72193 - 0$$

$$= 0,72193$$

Form Perhitungan IPM

Aplikasi Perhitungan Indeks Pembangunan Manusia

Angka Harapan Hidup: 2014

Angka Harapan Hidup: 69.89

Harapan Lama Sekolah: 12.72

Rata-Rata Lama Sekolah: 8.62

Pengeluaran Per Kapita: 10.891

Indeks Pendidikan: 0.2708

Indeks Pengeluaran Per Kapita: -4.8887

Indeks Pembangunan Manusia: 0.11812

Angka IPM dalam Ratusan: 11.818

Tahun	IPM	IPM Level	Kategori	Entropy	Gain
2010	0.11020	Low	Bad		
2011	0.11480	Medium	Good		
2012	0.11620	High	Good	0.72193	0.72193
2013	0.11746	High	Good		
2014	0.11812	High	Good		

Hasil Prediksi: 0,72193

Gambar 3.3 Hasil 3

3. IMPLEMENTASI PROGRAM

Perbandingan Hasil

Form Perhitungan IPM

Aplikasi Perhitungan Indeks Pembangunan Manusia

Angka Harapan Hidup: 2013

Angka Harapan Hidup: 69.77

Harapan Lama Sekolah: 12.51

Rata-Rata Lama Sekolah: 8.62

Pengeluaran Per Kapita: 10.824

Indeks Pendidikan: 0.2728

Indeks Pengeluaran Per Kapita: -4.8881

Indeks Pembangunan Manusia: 0.11629

Angka IPM dalam Ratusan: 11.629

Tahun	IPM	IPM Level	Kategori	Entropy	Gain
2010	0.11020	Low	Bad		
2011	0.11480	High	Good		
2012	0.11620	High	Good	0.9103	0.9103

Hasil Prediksi: 0,9103

Gambar 3.1 Hasil 1

Form Perhitungan IPM

Aplikasi Perhitungan Indeks Pembangunan Manusia

Angka Harapan Hidup: 2015

Angka Harapan Hidup: 70.90

Harapan Lama Sekolah: 12.88

Rata-Rata Lama Sekolah: 8.84

Pengeluaran Per Kapita: 10.718

Indeks Pendidikan: 0.2818

Indeks Pengeluaran Per Kapita: -4.8854

Indeks Pembangunan Manusia: 0.12121

Angka IPM dalam Ratusan: 12.121

Tahun	IPM	IPM Level	Kategori	Entropy	Gain
2010	0.11020	Low	Bad		
2011	0.11480	Medium	Bad		
2012	0.11620	Medium	Good		
2013	0.11746	Medium	Good	0.9103	0.9103
2014	0.11812	High	Good		
2015	0.12121	High	Good		

Hasil Prediksi: 0,9103

Gambar 3.4 Hasil 4

Form Perhitungan IPM

Aplikasi Perhitungan Indeks Pembangunan Manusia

Angka Harapan Hidup: 2013

Angka Harapan Hidup: 69.77

Harapan Lama Sekolah: 12.51

Rata-Rata Lama Sekolah: 8.62

Pengeluaran Per Kapita: 10.824

Indeks Pendidikan: 0.2749

Indeks Pengeluaran Per Kapita: -4.8874

Indeks Pembangunan Manusia: 0.11746

Angka IPM dalam Ratusan: 11.746

Tahun	IPM	IPM Level	Kategori	Entropy	Gain
2010	0.11020	Low	Bad		
2011	0.11480	Medium	Good		
2012	0.11620	High	Good	0.9103	0.9103
2013	0.11746	High	Good		

Hasil Prediksi: 0,9103

Gambar 3.2 Hasil 2

4. KESIMPULAN DAN SARAN

4.1 Kesimpulan

Secara keseluruhan, proses data mining hingga menghasilkan perhitungan algoritma pohon keputusan C4.5 yang dapat memberikan hasil prediksi untuk informasi kedepannya, maka dapat diambil beberapa kesimpulan sebagai berikut :

Berdasarkan dari pengolahan data yang penelitian lakukan dari 5 tahun terakhir 2010 sampai 2014 bisa dikatakan nilai IPM bagus dan nilai kategori tiap tahun naik berdasarkan pengujian yang dilakukan, nilai entropy IPM 0,72193 dan gain 0,72193 dimana nilai entropy didapatkan dari IPM level dan

kategori yang hasil nilai entropy untuk nilai penjumlahan nilai gain. Gain adalah nilai tertinggi dari pengolahan data IPM yang memiliki hasil untuk mendapatkan nilai prediksi.

4.2 Saran

Saran untuk penelitian selanjutnya :

1. Penelitian ini di data pada, angka harapan hidup, rata-rata lama sekolah, harapan lama sekolah, dan pengeluaran perkapita yang disesuaikan. Maka dari itu disarankan melakukan penelitian yang lebih luas dan hanya memprediksi data indek pembangunan manusia dan hasil nilai prediksi belum maksimal, maka dari itu dapat menggunakan perbandingan dua metode lain supaya dapat hasil yang memuaskan.
2. Penelitian sekarang prediksi pada IPM, untuk penelitian selanjutnya dapat dapat memprediksi angka harapan hidup, rata-rata lama sekolah, harapan lama sekolah, dan pengeluaran perkapita yang disesuaikan.
3. Prediksi IPM menggunakan data 5 tahun terakhir dari data yang didapat di BPS Kampar, disaran melakukan penelitian 6 tahun terakhir atau lebih dari 6 tahun.

DAFTAR PUSTAKA

- Badan Pusat Statistik (BPS). 2008. Indeks Pembangunan Manusia 2006 – 2007. BPS Kampar.
- Fajaryanto Adit Dwi; Alamsyah, A. (2016). Prediksi Indeks Pembangunan Manusia (IPM) Berdasarkan Jumlah Pengguna Internet, Produk Domestik Regional Bruto (PDRB), Jumlah Pekerja, Jumlah Penduduk Miskin Dan Populasi Penduduk Pada Tahun 2010
- Hastuti, K. (2012). Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif
- Kamagi, D. H., & Hansun, S. (2014). Implementasi Data Mining dengan Algoritma C4 . 5 untuk Memprediksi Tingkat Kelulusan Mahasiswa
- Kadir, Abdul (2003). Pengenalan Sistem Informasi, Andi Offset, Yogyakarta
- Mardi Yuli. (2014). Data Mining : Klasifikasi Menggunakan Algoritma C4.5
- Mabrur Angga Ginanjar & Lubis Riani (2012). Penerapan Data Mining Untuk memprediksi Kriteria Nasabah Kredit
- Pramadhani Aline Embun & Setiadi Tedy. (2014). Penerapan Data Mining Untuk Klasifikasi Prediksi Penyakit Ispa (Infeksi Saluran Pernapasan Akut) Dengan Algoritma Decision Tree (ID3)
- Pratama, A., & Al-Shaikh, M. (2012). Relation and Growth of Internet Penetration Rate with Human Development Level from 2000 to 2010
- Santoso Teguh Budi (2015). Analisa Dan Penerapan Metode C4.5 Untuk Prediksi Loyalitas Pelanggan
- Taslim, Fajrizal “Penerapan algoritma k-mean untuk clustering data obat pada puskesmas rumbai” (2016)