

Jawaban Topik 1 dan 2: Algoritma Decision Tree dan K-Means Clustering

Topik 1: Keputusan membeli produk

a. Algoritma Supervised Learning: Decision Tree

Untuk topik 1, saya akan menggunakan algoritma **Decision Tree** untuk memprediksi keputusan membeli produk berdasarkan atribut Umur, Penghasilan, dan Pekerjaan.

b. Simulasi Alur Kerja Algoritma Decision Tree

Langkah 1: Menghitung Entropi Total Dataset

Entropi mengukur ketidakpastian dalam dataset. Rumus entropi:

$$\text{Entropi}(S) = -\sum(p_i * \log_2(p_i))$$

Dimana p_i adalah proporsi kelas i dalam dataset.

Dalam dataset kita:

- Total instance: 10
- "Ya": 5 instances
- "Tidak": 5 instances

$$\text{Entropi}(S) = -[(5/10) * \log_2(5/10) + (5/10) * \log_2(5/10)]$$

$$\text{Entropi}(S) = -[0.5 * (-1) + 0.5 * (-1)]$$

$$\text{Entropi}(S) = -[-0.5 - 0.5]$$

$$\text{Entropi}(S) = 1$$

Langkah 2: Menghitung Information Gain untuk Setiap Atribut

Information Gain untuk Atribut "Umur"

Terdapat 3 nilai untuk Umur: Muda, Dewasa, dan Tua.

1. Menghitung entropi untuk "Muda":

- Total: 3 instances
- "Ya": 0 instances
- "Tidak": 3 instances

$$\text{Entropi(Muda)} = -[(0/3) * \log_2(0/3) + (3/3) * \log_2(3/3)] \text{ Entropi(Muda)} = -[0 + 1 * 0] \text{ Entropi(Muda)} = 0$$

2. Menghitung entropi untuk "Dewasa":

- Total: 4 instances
- "Ya": 4 instances
- "Tidak": 0 instances

$$\text{Entropi(Dewasa)} = -[(4/4) * \log_2(4/4) + (0/4) * \log_2(0/4)] \text{ Entropi(Dewasa)} = -[1 * 0 + 0]$$

$$\text{Entropi(Dewasa)} = 0$$

3. Menghitung entropi untuk "Tua":

- Total: 3 instances
- "Ya": 1 instance
- "Tidak": 2 instances

$$\begin{aligned} \text{Entropi(Tua)} &= -[(1/3) * \log_2(1/3) + (2/3) * \log_2(2/3)] \text{ Entropi(Tua)} = -[0.333 * (-1.585) + 0.667 * \\ &(-0.585)] \text{ Entropi(Tua)} = -[-0.528 - 0.39] \text{ Entropi(Tua)} = 0.918 \end{aligned}$$

Information Gain for Umur:

$$\text{IG(Umur)} = \text{Entropi(S)} - [(3/10) * \text{Entropi(Muda)} + (4/10) * \text{Entropi(Dewasa)} + (3/10) * \text{Entropi(Tua)}]$$

$$\text{IG(Umur)} = 1 - [0.3 * 0 + 0.4 * 0 + 0.3 * 0.918]$$

$$\text{IG(Umur)} = 1 - 0.275$$

$$\text{IG(Umur)} = 0.725$$

Information Gain untuk Atribut "Penghasilan"

Terdapat 3 nilai untuk Penghasilan: Rendah, Sedang, dan Tinggi.

1. Menghitung entropi untuk "Rendah":

- Total: 3 instances
- "Ya": 1 instance
- "Tidak": 2 instances

$$\text{Entropi(Rendah)} = -[(1/3) * \log_2(1/3) + (2/3) * \log_2(2/3)] \text{ Entropi(Rendah)} = 0.918$$

2. Menghitung entropi untuk "Sedang":

- Total: 5 instances
- "Ya": 3 instances
- "Tidak": 2 instances

$$\begin{aligned} \text{Entropi(Sedang)} &= -[(3/5) * \log_2(3/5) + (2/5) * \log_2(2/5)] \text{ Entropi(Sedang)} = -[0.6 * (-0.737) + 0.4 * \\ &(-1.322)] \text{ Entropi(Sedang)} = -[-0.442 - 0.529] \text{ Entropi(Sedang)} = 0.971 \end{aligned}$$

3. Menghitung entropi untuk "Tinggi":

- Total: 2 instances
- "Ya": 1 instance
- "Tidak": 1 instance

$$\text{Entropi(Tinggi)} = -[(1/2) * \log_2(1/2) + (1/2) * \log_2(1/2)] \quad \text{Entropi(Tinggi)} = -[0.5 * (-1) + 0.5 * (-1)]$$

$$\text{Entropi(Tinggi)} = 1$$

Information Gain for Penghasilan:

$$\text{IG(Penghasilan)} = \text{Entropi(S)} - [(3/10) * \text{Entropi(Rendah)} + (5/10) * \text{Entropi(Sedang)} + (2/10) * \text{Entropi(Tinggi)}]$$

$$\text{IG(Penghasilan)} = 1 - [0.3 * 0.918 + 0.5 * 0.971 + 0.2 * 1]$$

$$\text{IG(Penghasilan)} = 1 - [0.275 + 0.486 + 0.2]$$

$$\text{IG(Penghasilan)} = 1 - 0.961$$

$$\text{IG(Penghasilan)} = 0.039$$

Information Gain untuk Atribut "Pekerjaan"

Terdapat 3 nilai untuk Pekerjaan: Pelajar, Pekerja, dan Pensiunan.

1. Menghitung entropi untuk "Pelajar":

- Total: 3 instances
- "Ya": 1 instance
- "Tidak": 2 instances

$$\text{Entropi(Pelajar)} = -[(1/3) * \log_2(1/3) + (2/3) * \log_2(2/3)] \quad \text{Entropi(Pelajar)} = 0.918$$

2. Menghitung entropi untuk "Pekerja":

- Total: 5 instances
- "Ya": 4 instances
- "Tidak": 1 instance

$$\text{Entropi(Pekerja)} = -[(4/5) * \log_2(4/5) + (1/5) * \log_2(1/5)] \quad \text{Entropi(Pekerja)} = -[0.8 * (-0.322) + 0.2 * (-2.322)]$$

$$\text{Entropi(Pekerja)} = -[-0.258 - 0.464] \quad \text{Entropi(Pekerja)} = 0.722$$

3. Menghitung entropi untuk "Pensiunan":

- Total: 2 instances
- "Ya": 0 instances
- "Tidak": 2 instances

$$\text{Entropi(Pensiunan)} = -[(0/2) * \log_2(0/2) + (2/2) * \log_2(2/2)] \quad \text{Entropi(Pensiunan)} = 0$$

Information Gain for Pekerjaan:

$$\text{IG(Pekerjaan)} = \text{Entropi(S)} - [(3/10) * \text{Entropi(Pelajar)} + (5/10) * \text{Entropi(Pekerja)} + (2/10) * \text{Entropi(Pensiunan)}]$$

Entropi(Pensiunan)]

$$IG(Pekerjaan) = 1 - [0.3 * 0.918 + 0.5 * 0.722 + 0.2 * 0]$$

$$IG(Pekerjaan) = 1 - [0.275 + 0.361 + 0]$$

$$IG(Pekerjaan) = 1 - 0.636$$

$$IG(Pekerjaan) = 0.364$$

Langkah 3: Pemilihan Atribut Terbaik untuk Root Node

Berdasarkan perhitungan Information Gain:

- $IG(\text{Umur}) = 0.725$
- $IG(\text{Penghasilan}) = 0.039$
- $IG(\text{Pekerjaan}) = 0.364$

Karena $IG(\text{Umur})$ memiliki nilai tertinggi, maka "Umur" menjadi root node dalam decision tree.

Langkah 4: Pembagian Dataset dan Pembuatan Tree

Kita membagi dataset berdasarkan nilai atribut "Umur":

Untuk Umur = "Muda"

Semua instances memiliki label "Tidak", sehingga cabang ini langsung menjadi leaf node dengan keputusan "Tidak".

Untuk Umur = "Dewasa"

Semua instances memiliki label "Ya", sehingga cabang ini langsung menjadi leaf node dengan keputusan "Ya".

Untuk Umur = "Tua"

Masih terdapat ketidakpastian (1 "Ya" dan 2 "Tidak"), sehingga perlu dilakukan splitting lagi.

Untuk Umur = "Tua", terdapat 3 instances:

- (Tua, Sedang, Pekerja, Ya)
- (Tua, Tinggi, Pensiunan, Tidak)
- (Tua, Sedang, Pensiunan, Tidak)

Kita harus menghitung Information Gain untuk atribut "Penghasilan" dan "Pekerjaan" pada subset data ini.

Untuk simplisitas, karena hanya terdapat 3 instances, dan terlihat bahwa "Pekerjaan" sangat menentukan (semua "Pekerja" memiliki label "Ya" dan semua "Pensiunan" memiliki label "Tidak"), maka "Pekerjaan" akan menjadi node kedua untuk cabang "Tua".

Hasil Decision Tree:



Topik 2: Posisi lokasi pelanggan

a. Algoritma Unsupervised Learning: K-Means Clustering

Untuk topik 2, saya akan menggunakan algoritma **K-Means Clustering** untuk mengelompokkan pelanggan berdasarkan posisi lokasi mereka (koordinat X dan Y).

b. Simulasi Alur Kerja Algoritma K-Means Clustering

Langkah 1: Menentukan Jumlah Cluster (K)

Untuk dataset ini, saya akan menggunakan $K = 3$ (membagi pelanggan menjadi 3 kelompok).

Langkah 2: Memilih Centroid Awal

Kita akan memilih 3 titik acak dari dataset sebagai centroid awal:

- Centroid 1: (2, 5) [Titik 1]
- Centroid 2: (7, 5) [Titik 5]
- Centroid 3: (13, 8) [Titik 9]

Langkah 3: Iterasi 1 - Pengelompokkan Titik

Hitung jarak Euclidean dari setiap titik ke masing-masing centroid:

Contoh perhitungan untuk titik 1 (2, 5):

- Jarak ke Centroid 1 (2, 5): $\sqrt{[(2-2)^2 + (5-5)^2]} = 0$
- Jarak ke Centroid 2 (7, 5): $\sqrt{[(2-7)^2 + (5-5)^2]} = 5$
- Jarak ke Centroid 3 (13, 8): $\sqrt{[(2-13)^2 + (5-8)^2]} = \sqrt{(121 + 9)} = \sqrt{130} \approx 11.40$

Tabel Jarak dan Pengelompokkan:

No	Titik (X,Y)	Jarak ke C1	Jarak ke C2	Jarak ke C3	Cluster
1	(2, 5)	0	5	11.40	1
2	(3, 4)	1.41	4.12	11.18	1
3	(5, 8)	4.24	3.61	8.60	2
4	(6, 4)	4.12	1.41	9.22	2
5	(7, 5)	5	0	7.62	2
6	(8, 9)	7.21	4.12	5.83	2
7	(10, 5)	8	3	5	2
8	(11, 4)	9.22	4.12	5.10	3
9	(13, 8)	11.40	7.62	0	3
10	(14, 5)	12.04	7.07	4.24	3

Langkah 4: Iterasi 1 - Menghitung Centroid Baru

Cluster 1:

- Titik: (2, 5), (3, 4)
- Centroid Baru C1: $((2+3)/2, (5+4)/2) = (2.5, 4.5)$

Cluster 2:

- Titik: (5, 8), (6, 4), (7, 5), (8, 9), (10, 5)
- Centroid Baru C2: $((5+6+7+8+10)/5, (8+4+5+9+5)/5) = (7.2, 6.2)$

Cluster 3:

- Titik: (11, 4), (13, 8), (14, 5)
- Centroid Baru C3: $((11+13+14)/3, (4+8+5)/3) = (12.67, 5.67)$

Langkah 5: Iterasi 2 - Pengelompokkan Titik Dengan Centroid Baru

Kembali menghitung jarak Euclidean dari setiap titik ke masing-masing centroid baru:

Tabel Jarak dan Pengelompokkan Iterasi 2:

No	Titik (X,Y)	Jarak ke C1	Jarak ke C2	Jarak ke C3	Cluster
1	(2, 5)	0.71	5.27	10.86	1
2	(3, 4)	0.71	4.34	9.90	1
3	(5, 8)	5.00	2.82	8.24	2
4	(6, 4)	3.54	2.28	6.76	2
5	(7, 5)	4.53	1.20	5.67	2
6	(8, 9)	7.28	2.82	6.06	2
7	(10, 5)	7.52	2.82	2.70	3
8	(11, 4)	8.56	4.19	1.79	3
9	(13, 8)	11.31	6.36	2.50	3
10	(14, 5)	11.58	7.04	1.99	3

Langkah 6: Iterasi 2 - Menghitung Centroid Baru

Cluster 1:

- Titik: (2, 5), (3, 4)
- Centroid Baru C1: (2.5, 4.5) [Tidak berubah dari iterasi sebelumnya]

Cluster 2:

- Titik: (5, 8), (6, 4), (7, 5), (8, 9)
- Centroid Baru C2: $((5+6+7+8)/4, (8+4+5+9)/4) = (6.5, 6.5)$

Cluster 3:

- Titik: (10, 5), (11, 4), (13, 8), (14, 5)
- Centroid Baru C3: $((10+11+13+14)/4, (5+4+8+5)/4) = (12, 5.5)$

Langkah 7: Melanjutkan Iterasi Hingga Konvergen

Kita terus mengulang langkah-langkah di atas hingga centroid tidak berubah lagi atau perubahan sangat kecil.

Untuk kesederhanaan, kita asumsikan setelah beberapa iterasi, kita mendapatkan hasil final:

Hasil Clustering Final:

Cluster 1:

- Titik: (2, 5), (3, 4)

- Centroid Final: (2.5, 4.5)
- Interpretasi: Pelanggan di lokasi bagian barat

Cluster 2:

- Titik: (5, 8), (6, 4), (7, 5), (8, 9)
- Centroid Final: (6.5, 6.5)
- Interpretasi: Pelanggan di lokasi bagian tengah

Cluster 3:

- Titik: (10, 5), (11, 4), (13, 8), (14, 5)
- Centroid Final: (12, 5.5)
- Interpretasi: Pelanggan di lokasi bagian timur

Visualisasi Hasil Clustering:



Keterangan:

Cluster 1: (2,5), (3,4)

Cluster 2: (5,8), (6,4), (7,5), (8,9)

Cluster 3: (10,5), (11,4), (13,8), (14,5)

Kesimpulan:

K-Means clustering berhasil mengelompokkan 10 titik lokasi pelanggan menjadi 3 cluster berdasarkan

kedekatan geografis mereka.