

Korelasi Minat Anak Keluarga Petani dan Non Petani Dalam Melanjutkan Pendidikan Ke Perguruan Tinggi Terhadap Faktor Penyebab Penghasilan Dengan Metode Apriori (Studi Kasus: Desa Pasar VIII Namo Terasi)

The Correlation of Interests of Children of Farmer and Non-Farming Families in Continuing Education to Higher Education on Factors That Cause Income Using the Apriori Method (Case Study: Desa Pasar VIII Namo Terasi)

Karonika Br Barus^{*1}, Novriyenni², Magdalena Simanjuntak³

^{1,2,3}Prodi Sistem Informasi STMIK KAPUTAMA, Jl. Veteran No.4A-9A, Binjai, Sumatera Utara, 20714, Indonesia

e-mail: 1karowbarus@gmail.com, 2novriyenni.sikumbang@gmail.com,
3magdalena.simanjuntak84@gmail.com

Abstrak

Berkurangnya minat belajar seorang individu disebabkan karena berbagai faktor. Salah satunya adalah faktor sosial dimana dalam beberapa kondisi menyebabkan individu merasa rendah diri untuk melanjutkan pendidikan ke ranah akademis yang lebih tinggi. Bagi sebagian orang profesi petani kerap dianggap remeh menyebabkan nilai sosial profesi petani menurun sehingga menyebabkan generasi petani berikutnya merasa tidak percaya diri untuk melanjutkan pendidikan ke tahap yang lebih tinggi. Dari hasil pengujian pada 24 sampel data desa yang ada di kantor Desa Pasar VIII Namo Terasi menggunakan metode apriori, ditemukan sebanyak 1 rule yaitu Jika Pekerjaan orang tua adalah Pegawai negeri dan penghasilan diatas 4 juta, maka anak akan kuliah dengan nilai pendukung sebesar 20.8%, nilai kepastian sebesar 50%.

Kata kunci: Data Mining, Itemset, Metode Apriori, Rule

Abstract

The decrease in an individual's interest in learning is caused by various factors. One of them is social factors which in some conditions cause individuals to feel inferior to continue their education to a higher academic realm. For some people, the farming profession is often underestimated, causing the social value of the farming profession to decline, causing the next generation of farmers to feel insecure to continue their education to a higher stage. From the test results on 24 samples of village data in the Pasar VIII Namo Terasi village office using the a priori method, it was found that there was 1 rule, namely If the parent's occupation is a civil servant and the income is above 4 million, the child will go to college with a supporting value of 20.8%, confidence value of 50%.

Keywords: Data Mining, Itemset, Apriori Method, Rule

1. PENDAHULUAN

Berkurangnya minat belajar seorang individu disebabkan karena berbagai faktor. Salah satunya adalah faktor sosial dimana dalam beberapa kondisi menyebabkan individu merasa rendah diri untuk melanjutkan pendidikan ke ranah akademis yang lebih tinggi. Bagi sebagian orang profesi petani kerap dianggap remeh menyebabkan nilai sosial profesi petani menurun sehingga menyebabkan generasi petani berikutnya merasa tidak percaya diri untuk melanjutkan pendidikan ke tahap yang lebih tinggi.

Pada Desa Pasar VIII Namo Terasi telah dilakukan survei dengan metode wawancara dan studi kepustakaan sehingga dapat dihasilkan Data Desa Pasar VIII Namo terasi yang dapat diolah dengan metode apriori. Metode apriori adalah pengambilan data dengan aturan asosiasi (Association rule) untuk menentukan hubungan asosiatif suatu kombinasi item. Association rule yang dimaksud dilakukan melalui mekanisme perhitungan support dan confidence dari suatu hubungan item. Sebuah rule asosiasi dikatakan interesting jika nilai support adalah lebih besar dari minimum support dan juga nilai confidence adalah lebih besar dari minimum confidence. Metode apriori ini akan cocok untuk diterapkan bila terdapat beberapa hubungan item yang ingin dianalisis.

2. METODE PENELITIAN

Korelasi Menurut (Kusrini, Emha Taufiq Lutfhi, 2009, h. 149) Algoritma A priori termasuk jenis aturan asosiasi pada data mining. Selain apriori, yang termasuk pada golongan ini adalah metode generalize Rule Indication dan Algoritma Hans Based. Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antar suatu kombinasi item.

Analisis asosiasi juga dikenal sebagai salah satu teknik data mining yang menjadi dasar dari berbagai teknik data mining lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut dengan analisis pola frekuensi tinggi (frequent pattern mining) menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, yaitu support (nilai penunjang) adalah persentase kombinasi item tersebut dalam database dan confidence (nilai kepastian) adalah kuatnya hubungan antar item dalam suatu aturan asosiatif.

3. HASIL DAN PEMBAHASAN

Dari hasil ranking perhitungan metode MOORA diatas, berdasarkan nilai tertinggi maka industri kecil menengah yang direkomendasikan menjadi produk unggulan daerah adalah Industri Anyaman Bambu.

Metodologi dasar Analisis Asosiasi terbagi menjadi dua tahapan yaitu:

1. Analisis Pola Frekuensi Tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam *database*.

2. Pembentukan Aturan Asosiasi

Setelah semua frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi *confidence* dengan menghitung aturan asosiatif.

Untuk mencari korelasi atau hubungan pada data faktor penyebab penghasilan penduduk di di kantor Desa Pasar VIII Namo Terasi, penulis mengambil 25 data sebagai berikut :

Data Desa Pasar VIII Namo Terasi

No	Pekerjaan	Penghasilan	Status Anak	Pendidikan
1	PNS	RP.5.000.000		KULIAH
2	PETANI	RP.7.000.000		KULIAH
3	PETANI	RP.4.000.000		KULIAH
4	WIRASWASTA	RP.3.500.000		KULIAH
5	WIRASWASTA	RP.6.000.000		KULIAH
6	PETANI	RP.3.500.000		KULIAH
7	WIRASWASTA	RP.12.000.000		TIDAK
8	PETANI	RP.6.000.000		TIDAK
9	PNS	RP.6.500.000		KULIAH
10	PETANI	RP.8.000.000		TIDAK
11	PETANI	RP.5.000.000		TIDAK
12	PETANI	RP.5.000.000		KULIAH
13	PETANI	RP.3.500.000		KULIAH
14	PETANI	RP.4.000.000		KULIAH
15	PETANI	RP.3.500.000		KULIAH
16	PETANI	RP.4.000.000		TIDAK
17	PNS	RP.5.500.000		KULIAH
18	PNS	RP.7.000.000		KULIAH
19	PETANI	RP.3.500.000		TIDAK
20	PETANI	RP.4.000.000		KULIAH
21	PNS	RP.5.500.000		KULIAH
22	PETANI	RP.6.000.000		TIDAK
23	PNS	RP.5.000.000		KULIAH
24	PETANI	RP.5.500.000		KULIAH

Representasi Data Pekerjaan

Pekerjaan	Kode
Petani	PE
Karyawan	KR
Wiraswasta	WS
PNS	PN

Representasi Data Penghasilan

Penghasilan	Kode
< Rp. 999.000	A0
>= Rp. 1.000.000 & <=1.999.000	A1
>= Rp. 2.000.000 & <= 2.999.000	A2
>= Rp. 3.000.000 & <= 3.999.000	A3
>= Rp. 4.000.000	A4

Data Status Pendidikan Anak

Status Pendidikan Anak	Kode
Kuliah	K
Tidak Kuliah	T

Representasi Data Desa Pasar VIII Namo Terasi

No	Pekerjaan				Penghasilan					Status Pendidikan Anak	
	PE	KA	WU	PN	A0	A1	A2	A3	A4	K	T
1	0	0	0	1	0	0	0	0	1	1	0
2	1	0	0	0	0	0	0	0	1	1	0
3	1	0	0	0	0	0	0	0	1	1	0
4	0	0	1	0	0	0	0	1	0	1	0
5	0	0	1	0	0	0	0	0	1	1	0
6	1	0	0	0	0	0	0	1	0	1	0
7	0	0	1	0	0	0	0	0	1	0	1
8	1	0	0	0	0	0	0	0	1	0	1
9	0	0	0	1	0	0	0	0	1	1	0
10	1	0	0	0	0	0	0	0	1	0	1
11	1	0	0	0	0	0	0	0	1	0	1
12	1	0	0	0	0	0	0	0	1	1	0
13	1	0	0	0	0	0	0	1	0	1	0
14	1	0	0	0	0	0	0	0	1	1	0

15	1	0	0	0	0	0	0	1	0	1	0
16	1	0	0	0	0	0	0	0	1	0	1
17	0	0	0	1	0	0	0	0	1	1	0
18	0	0	0	1	0	0	0	0	1	1	0
19	1	0	0	0	0	0	0	1	0	0	1
20	1	0	0	0	0	0	0	0	1	1	0
21	0	0	0	1	0	0	0	0	1	1	0
22	1	0	0	0	0	0	0	0	1	0	1
23	0	0	0	1	0	0	0	0	1	1	0
24	1	0	0	0	0	0	0	0	1	1	0
Jlh	15	0	3	6	0	0	0	5	19	17	7

3. Tentukan θ (Frequent)

Misalkan kita tentukan $\theta > 3$, maka kita dapat menentukan *frequent itemset*. Dari tabel diatas diketahui total θ untuk $k = 1$, ada beberapa yang lebih besar dari θ .

4. Tentukan *Item Set*

Maka F1 untuk tabel Data Desa Pasar VIII Namo Terasi yang nilainya lebih besar dari θ yaitu semua variable dapt digunakan dalam perhitungan karena lebih besar dari 3.

5. Ujikan Set θ

Himpunan yang mungkin terbentuk {PE,A3,K}, {PN,A3,K}, {PE,A4,T}

Calon 3 Itemset

K	PE	A3	K	f
1	0	0	1	N
2	1	0	1	N
3	1	0	1	N
4	0	1	1	N
5	0	0	1	N
6	1	1	1	Y
7	0	0	0	N
8	1	0	0	N
9	0	0	1	N
10	1	0	0	N
11	1	0	0	N
12	1	0	1	N

K	PN	A3	K	f
1	1	0	1	N
2	0	0	1	N
3	0	0	1	N
4	0	1	1	N
5	0	0	1	N
6	0	1	1	N
7	0	0	0	N
8	0	0	0	N
9	1	0	1	N
10	0	0	0	N
11	0	0	0	N
12	0	0	1	N

K	PE	A4	T	f
1	0	1	0	N
2	1	1	0	N
3	1	1	0	N
4	0	0	0	N
5	0	1	0	N
6	1	0	0	N
7	0	1	1	N
8	1	1	1	Y
9	0	1	0	N
10	1	1	1	Y
11	1	1	1	Y
12	1	1	0	N

13	1	1	1	Y	13	0	1	1	N	13	1	0	0	N
14	1	0	1	N	14	0	0	1	N	14	1	1	0	N
15	1	1	1	Y	15	0	1	1	N	15	1	0	0	N
16	1	0	0	N	16	0	0	0	N	16	1	1	1	Y
17	0	0	1	N	17	1	0	1	N	17	0	1	0	N
18	0	0	1	N	18	1	0	1	N	18	0	1	0	N
19	1	1	0	N	19	0	1	0	N	19	1	0	1	N
20	1	0	1	N	20	0	0	1	N	20	1	1	0	N
21	0	0	1	N	21	1	0	1	N	21	0	1	0	N
22	1	0	0	N	22	0	0	0	N	22	1	1	1	Y
23	0	0	1	N	23	1	0	1	N	23	0	1	0	N
24	1	0	1	Y	24	0	0	1	N	24	1	1	0	N
Jumlah				3	Jumlah				0	Jumlah				5

Dari tabel 3 unsur diatas, Y artinya item yang saling berhubungan, sedangkan N berarti tidak ada item yang saling berhubungan Jumlah frekuensi *itemset* harus lebih besar dengan jumlah frekuensi *itemset*. Dari tabel diatas maka di dapat f_3 : {PE,A3,K}, {PE,A4,T}

Untuk mengetahui hubungan atau korelasi antar item kekuatan hubungan ditentukan oleh 2 faktor yaitu *support* dan *confidence*, yang diperoleh dengan rumus sebagai berikut:

Support = Nilai pendukung

$$Support = \frac{\sum \text{Item yang digunakan}}{\sum \text{Jumlah seluruh transaksi}} \times 100\%$$

Confidence = Nilai Kepastian

$$Confidence = \frac{\sum \text{Item yang digunakan sekaligus}}{\sum \text{Jumlah transaksi pada bagian antecedent}} \times 100\%$$

Rule untuk 3 *itemset* terdiri atas 2 rule yang di implementasikan pada tabel berikut :

Aturan Asosiasi 3 *Itemset* S * C

If antecedent then consequent	Support	Confidence	S*C
If Pekerjaan PE dan Penghasilan A3 → Kuliah K	12.5%	75%	9.37 %
If Pekerjaan PN dan Penghasilan A4 → Kuliah K	20.8%	50%	10.4 %

Setelah melakukan perkalian antara support dan *confidence*, maka hasil dari perkalian tersebut maka S*C yang terbesar yang akan menjadi *Best Rule*.

Best Rule

If antecedent then consequent	Support	Confidence	S*C
If Pekerjaan PN dan Penghasilan A4 → Kuliah K	20.8%	50%	10.4 %

DAFTAR PUSTAKA

- [1] Buaton, Relita. 2013. Mudahnya Membuat Website (HTML, CSS, Javascript, PHP). Edisi pertama, Andi, Yogyakarta.
- [2] Budi Sutedjo, S. M., & Michael AN, S. 2000. Algoritma & Teknik Pemrograman. Andi, Yogyakarta.
- [3] Jonathan, Sarwono. 2006. Metode Penelitian Kuantitatif dan Kualitatif. Graha Ilmu, Yogyakarta.
- [4] Larose D, T., 2005, Discovering knowledge in data : an introduction to data mining, Jhon Wiley & Sons Inc.
- [5] Luthfi emha taufiq dan Kusri. 2009. Algoritma Data Mining. Edisi Pertama, Andi, Yogyakarta.
- [6] Nugroho, Bunafit. 2005. Database Relasional dengan MySQL. Edisi Pertama Andi, Yogyakarta.
- [7] Prasetyo, Eko. 2014. Data Mining. Andi Offset, Yogyakarta.
- [8] Santosa, Budi. 2007. Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Graha Ilmu, Yogyakarta.
- [9] Sugiarti Yuni. 2013. Analisis Dan Perancangan UML (Unified Modeling Language) Generated VB.6. Edisi Pertama, Graha Ilmu, Yogyakarta.
- [10] <https://core.ac.uk/download/pdf/12351324.pdf> diakses tanggal 16 Mei 2022.