

PERBANDINGAN ROUGH SET DAN ALGORITMA APRIORI UNTUK SISTEM REKOMENDASI PERPUSTAKAAN

Jaka Nugraha¹, Muhammad Muhajir², Rachmad Febrian³

¹Universitas Islam Indonesia, jnugraha@uii.ac.id

²Universitas Islam Indonesia, mmuhajir@uii.ac.id

³Universitas Islam Indonesia, febrian.rachmad.rf@gmail.com

Abstract. The recommendation system is a dynamic information filtering system that is produced according to user interests or behavior. The recommendation system in the library can provide book references based on user interests or characteristics. UII Central Library has an information system in the form of a database of book lending transactions that can be used for a recommendation system. The method is a rough set and apriori algorithm. This study compares 2 methods to get the best method that can be applied in the recommendation system. The results obtained by the number of rough set rules as many as 14 rules with an average coverage of 0,01111 and average accuracy of 0,87416 and the number of apriori algorithm rules as many as 23 rules with an average support of 0,00276 and average confidence of 0,87458. Based on the number rules, the average value of accuracy or confidence, the apriori algorithm method is a method that can be used for the recommendation system in the UII Central Library.

Keywords: Recommendation System, Rough Set, Apriori Algorithm.

Abstrak. Sistem rekomendasi adalah sistem penyaringan informasi secara dinamis yang dihasilkan sesuai dengan minat atau perilaku pengguna. Sistem rekomendasi pada perpustakaan dapat memberikan referensi buku berdasarkan minat atau karakteristik pengguna. Perpustakaan Pusat UII memiliki sistem informasi berupa database transaksi peminjaman buku yang dapat digunakan untuk sistem rekomendasi. Untuk menemukan pola yang akan dibuat sistem rekomendasi diperlukan sebuah metode analisis. Metode analisis yang digunakan yaitu roughset dan algoritma apriori. Penelitian ini membandingkan 2 metode untuk mendapatkan metode terbaik yang dapat diterapkan dalam sistem rekomendasi. Hasil penelitian diperoleh jumlah aturan rough set sebanyak 14 aturan dengan rata-rata *coverage* sebesar 0,01111 serta rata-rata *accuracy* sebesar 0,87416 dan jumlah aturan algoritma apriori sebanyak 23 aturan dengan rata-rata *support* sebesar 0,00276 dan rata-rata *confidence* sebesar 0,87458. Berdasarkan jumlah aturan, nilai rata-rata *coverage* atau *support* serta nilai rata-rata *accuracy* atau *confidence* didapatkan metode algoritma apriori adalah metode yang dapat digunakan untuk sistem rekomendasi di Perpustakaan Pusat UII.

Kata Kunci: Sistem Rekomendasi, Rough Set, Algoritma Apriori.

1 Pendahuluan

Sistem rekomendasi adalah sistem penyaringan informasi secara dinamis yang dihasilkan sesuai dengan minat atau perilaku pengguna [1]. Sistem rekomendasi tidak hanya dibutuhkan dalam bidang bisnis tetapi juga dibutuhkan pada sistem

perpustakaan. Sistem rekomendasi pada perpustakaan dapat memberikan referensi buku berdasarkan minat atau karakteristik pengguna perpustakaan.

Perpustakaan Pusat UII memiliki sistem informasi database transaksi peminjaman buku, tetapi mengalami penumpukan data yang jumlahnya besar, padahal data tersebut dapat digunakan untuk sistem rekomendasi. Sistem rekomendasi dibuat dari data history transaksi peminjaman buku yang berupa pola-pola peminjaman buku. Pola-pola tersebut didapatkan dari hasil teknik data mining. Data mining merupakan proses untuk menggali pengetahuan-pengetahuan yang tersembunyi didalam database sehingga menjadi informasi yang berharga [2]. Terdapat beberapa metode dalam data mining yang sering digunakan, yaitu *rough set* dan algoritma apriori.

Rough set merupakan metode atau teknik yang efisien dalam penambangan basis data atau penemuan pengetahuan dalam basis data relasional. *Rough set* digunakan dalam berbagai masalah, seperti representasi pengetahuan yang tidak pasti atau tidak tepat, analisis pengetahuan, evaluasi kualitas dan ketersediaan informasi yang berkenaan dengan konsistensi [3]. Kelebihan metode *rough set* yaitu menyediakan algoritma yang efisien untuk menemukan pola yang tersembunyi dalam data sehingga cepat dan mudah [4]. Sedangkan algoritma apriori merupakan metode yang digunakan untuk menemukan pola dalam sebuah data [3]. Dengan mengeliminasi kandidat rule atau aturan berdasarkan nilai support dan confidence membuat algoritma apriori memerlukan waktu yang lebih sedikit dan lebih cepat dalam proses komputasi [5].

Berdasarkan uraian diatas, pada makalah ini dibahas perbandingan metode *rough set* dan algoritma apriori untuk mendapatkan metode terbaik yang dapat diterapkan dalam sistem rekomendasi.

2 Kajian Teori

2.1 Rough Set

Rough set merupakan metode atau teknik yang efisien dalam penambangan basis data atau penemuan pengetahuan dalam basis data relasional. *Rough set* digunakan dalam berbagai masalah, seperti representasi pengetahuan yang tidak pasti atau tidak tepat, analisis pengetahuan, evaluasi kualitas dan ketersediaan informasi yang berkenaan dengan konsistensi [3]. Kelebihan metode *rough set* yaitu menyediakan algoritma yang efisien untuk menemukan pola yang tersembunyi dalam data sehingga cepat dan mudah [4]. Validasi atau evaluasi performa *rough set* dilihat dari 2 nilai, yaitu *coverage* dan *accuracy*.

2.1.1 Coverage

Coverage merupakan nilai rasio perbandingan dari objek yang memenuhi *antecedent* dan juga memenuhi *conclusion* terhadap objek yang memenuhi *conclusion* saja.

$$Coverage (M \Rightarrow N) = \frac{support (M \Rightarrow N)}{card (||g||_I)} \quad (1)$$

2.1.2 Accuracy

Accuracy merupakan nilai rasio perbandingan dari objek yang memenuhi *antecedent* dan juga memenuhi *conclusion* terhadap objek yang memenuhi *antecedent* saja.

$$Accuracy (M \Rightarrow N) = \frac{support (M \Rightarrow N)}{card (||f||_I)} \quad (2)$$

2.2 Algoritma Apriori

Algoritma apriori merupakan salah satu algoritma dalam *association rules*. Algoritma apriori digunakan untuk menemukan pola dalam sebuah data [6]. Dengan mengeliminasi kandidat rule atau aturan berdasarkan nilai support dan confidence membuat algoritma apriori memerlukan waktu yang lebih sedikit dan lebih cepat dalam proses komputasi[5]. Validasi dalam algoritma apriori dilihat dari 2 nilai, yaitu nilai *support* dan nilai *confidence*.

2.2.1 Support

Support merupakan nilai persentase dari transaksi yang berisi itemset dari keseluruhan transaksi atau seberapa sering itemset muncul di dataset. Berikut adalah rumus untuk mencari nilai *support* [7]:

$$\text{Support } (M \rightarrow N) = P(M \cup N) = \frac{n(M \cup N)}{n(S)} \quad (3)$$

Dimana:

$P(M \cup N)$ = Probabilitas kejadian M dan N secara bersamaan

$n(M \cup N)$ = Banyaknya kejadian M dan N secara bersamaan

$n(S)$ = Banyaknya anggota S

2.2.2 Confidence

Confidence merupakan nilai keyakinan yang mengindikasikan seberapa sering aturan itu terbukti benar atau perbandingan antara nilai support dari himpunan item. Nilai *confidence* sering disebut *accuracy*.

$$\text{Confidence } (M \rightarrow N) = P(N|M) = \frac{P(M \cup N)}{P(M)} \quad (4)$$

Dimana :

$P(N|M)$ = Probabilitas bersyarat dari kejadian N bila kejadian M telah terjadi

$P(M \cup N)$ = Probabilitas kejadian M dan N secara bersamaan

$P(M)$ = Probabilitas kejadian M

3 Metode Penelitian

Metode yang digunakan adalah rough set dan algoritma apriori. Penelitian ini membandingkan rough set dan algoritma apriori pada data peminjaman buku di Perpustakaan Pusat UII dimana nantinya hasil analisis ini untuk dibuat sebuah sistem rekomendasi buku. Variabel yang digunakan yaitu variabel Angkatan yang terdiri dari tahun 2010 s/d 2017, variabel Prodi untuk jenjang sarjana sebanyak 24 program studi dan variabel Kategori adalah kategori buku yang dikelompokkan menjadi 10 kategori. Data yang digunakan sebanyak 55.982 data.

4 Hasil dan Pembahasan

4.1 Rough Set

Tabel 1 menyajikan hasil aturan yang terbentuk dari *rough set* :

Tabel 1 Hasil Aturan dari *Rough Set*

No.	Aturan Asosiasi	Coverage	Accuracy
-----	-----------------	----------	----------

No.	Aturan Asosiasi	Coverage	Accuracy
1	ANGKATAN(ANGKATAN 2013) AND PRODI(ILMU EKONOMI) => KATEGORI(ILMU-ILMU SOSIAL)	0.003506	0.829787
2	ANGKATAN(ANGKATAN 2011) AND PRODI(TEKNIK SIPIL) =>KATEGORI(ILMU TERAPAN (TEKNOLOG))	0.0084	0.905983
3	ANGKATAN(ANGKATAN 2012) AND PRODI(TEKNIK SIPIL) => KATEGORI(ILMU-ILMU MURNI)	0.041465	0.843296
4	ANGKATAN(ANGKATAN 2010) AND PRODI(TEKNIK SIPIL) => KATEGORI(ILMU TERAPAN (TEKNOLOGI))	0.002383	0.909091
5	ANGKATAN(ANGKATAN 2014) AND PRODI(TEKNIK MESIN) => KATEGORI(ILMU TERAPAN (TEKNOLOGI))	0.001748	0.846154
6	ANGKATAN(ANGKATAN 2011) AND PRODI(MANAJEMEN) => KATEGORI(ILMU TERAPAN (TEKNOLOGI))	0.002304	0.935484
7	ANGKATAN(ANGKATAN 2010) AND PRODI(EKONOMI ISLAM) => KATEGORI(AGAMA)	0.000426	0.857143
8	ANGKATAN(ANGKATAN 2012) AND PRODI(FARMASI) => KATEGORI(ILMU TERAPAN (TEKNOLOGI))	0.005402	0.894737
9	ANGKATAN(ANGKATAN 2013) AND PRODI(TEKNIK MESIN) => KATEGORI(ILMU TERAPAN (TEKNOLOGI))	0.002224	0.903226
10	ANGKATAN(ANGKATAN 2010) AND PRODI(TEKNIK INDUSTRI) => KATEGORI(ILMU TERAPAN (TEKNOLOGI))	0.002621	0.891892
11	ANGKATAN(ANGKATAN 2011) AND PRODI(HUKUM) => KATEGORI(ILMU-ILMU SOSIAL)	0.001798	0.952381
12	ANGKATAN(ANGKATAN 2012) AND PRODI(KIMIA) => KATEGORI(ILMU-ILMU MURNI)	0.001061	0.800000
13	ANGKATAN(ANGKATAN 2017) AND PRODI(HUKUM ISLAM) => KATEGORI(AGAMA)	0.064594	0.859301
14	ANGKATAN(ANGKATAN 2017) AND PRODI(TEKNIK MESIN) => KATEGORI(ILMU-ILMU MURNI)	0.017511	0.809816

Dari **Tabel 1** menggunakan batasan nilai *coverage* sebesar 0.0001 dan batasan nilai *accuracy* sebesar 0.8 didapatkan sebanyak 14 aturan. Aturan pertama dengan nilai *coverage* sebesar 0.003506 yang artinya sebanyak 0.35% transaksi terjadi dengan aturan asosiasi “jika angkatan 2013 dan dari program studi ilmu ekonomi maka buku yang dipinjam adalah buku dengan kategori ilmu-ilmu sosial dan mempunyai nilai *accuracy* sebesar 0.82. Begitu seterusnya untuk aturan yang lainnya. Rata-rata *coverage* sebesar 0.01111 dan rata-rata *accuracy* sebesar 0.87416.

4.2 Algoritma Apriori

Hasil aturan yang terbentuk dari algoritma apriori selengkapnya disajikan pada Tabel 2.:

Tabel 2. Hasil Aturandari Algoritma Apriori

No.	Aturan Asosiasi	Support	Confidence
1	ANGKATAN=TAHUN 2010, PRODI=HUKUM =>KATEGORI=AGAMA	0.0008	0.81
2	ANGKATAN=TAHUN 2010, PRODI=TEKNIK INDUSTRI => KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.0006	0.89
3	ANGKATAN=TAHUN 2010, PRODI=TEKNIK SIPIL =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.0005	0.91
4	ANGKATAN=TAHUN 2010, PRODI=EKONOMI ISLAM =>KATEGORI=AGAMA	0.0001	0.86
5	ANGKATAN=TAHUN 2013, PRODI=ILMU EKONOMI =>KATEGORI=ILMU-ILMU SOSIAL	0.0007	0.83
6	ANGKATAN=TAHUN 2011, PRODI=HUKUM =>KATEGORI=ILMU-ILMU SOSIAL	0.0004	0.95
7	ANGKATAN=TAHUN 2011, PRODI=MANAJEMEN =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.0005	0.94
8	ANGKATAN=TAHUN 2011, PRODI=TEKNIK SIPIL =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.002	0.91
9	ANGKATAN=TAHUN 2017, PRODI=TEKNIK MESIN =>KATEGORI=ILMU-ILMU MURNI	0.002	0.81
10	ANGKATAN=TAHUN 2013, PRODI=TEKNIK MESIN =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.0005	0.90
11	ANGKATAN=TAHUN 2015, PRODI=TEKNIK MESIN =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.0003	1

No.	Aturan Asosiasi	Support	Confidence
12	ANGKATAN=TAHUN 2014, PRODI=TEKNIK MESIN =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.0004	0.85
13	ANGKATAN=TAHUN 2012, PRODI=MANAJEMEN =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.0003	0.94
14	ANGKATAN=TAHUN 2016, PRODI=MANAJEMEN =>KATEGORI=AGAMA	0.001	0.85
15	ANGKATAN=TAHUN 2014, PRODI=PENDIDIKAN DOKTER=>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.001	0.82
16	ANGKATAN=TAHUN 2015, PRODI=ARSITEKTUR => KATEGORI=KESENIAN/OLAHRAGA	0.002	0.80
17	ANGKATAN=TAHUN 2012, PRODI=FARMASI => KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.001	0.89
18	ANGKATAN=TAHUN 2012, PRODI=KIMIA =>KATEGORI=ILMU-ILMU MURNI	0.0001	0.80
19	ANGKATAN=TAHUN 2012, PRODI=TEKNIK KIMIA =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.0006	0.93
20	ANGKATAN=TAHUN 2012, PRODI=TEKNIK SIPIL =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.009	0.84
21	ANGKATAN=TAHUN 2013, PRODI=TEKNIK SIPIL =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.001	0.87
22	ANGKATAN=TAHUN 2014, PRODI=TEKNIK SIPIL =>KATEGORI=ILMU TERAPAN (TEKNOLOGI)	0.01	0.84
23	ANGKATAN=TAHUN 2017, PRODI=HUKUM ISLAM =>KATEGORI=AGAMA	0.01	0.86

Berdasarkan **Tabel 2** menggunakan batasan nilai *support* sebesar 0.0001 dan batasan nilai *confidence* sebesar 0.8 didapatkan sebanyak 23 aturan. Aturan pertama dengan nilai *coverage* sebesar 0.0008 yang artinya sebanyak 0.08% transaksi terjadi dengan aturan asosiasi “jika angkatan 2010 dan dari program studi hukum maka buku yang dipinjam adalah buku dengan kategori agama serta nilai *accuracy* sebesar 0.81. Begitu seterusnya untuk aturan yang lainnya. Rata-rata *support* sebesar 0.00276 dan rata-rata *confidence* sebesar 0.87458.

5 Kesimpulan

Berdasarkan hasil aturan atau *rule* yang didapatkan, jumlah aturan rough set sebanyak 14 aturan dengan rata-rata *coverage* sebesar 0.01111 serta rata-rata

accuracy sebesar 0.87416. Jumlah aturan algoritma apriori sebanyak 23 aturan dengan rata-rata *support* sebesar 0.00276 dan rata-rata *confidence* sebesar 0.87458. Berdasarkan jumlah aturan, nilai rata-rata coverage atau support serta nilai rata-rata *accuracy* atau *confidence* didapatkan metode algoritma apriori adalah metode yang dapat digunakan untuk sistem rekomendasi di Perpustakaan Pusat UII.

Penelitian ini dapat dikembangkan menggunakan metode atau algoritma lainnya untuk menemukan pola-pola atau aturan untuk dilakukan perbandingan.

6 Ucapan Terima Kasih

Peneliti mengucapkan terimakasih kepada Direktorat Penelitian dan Pengabdian Masyarakat Universitas Islam Indonesia (DPPM UII) yang telah mendukung sepenuhnya penelitian ini untuk skema tahap kedua tahun 2018/2019.

Daftar Pustaka

- [1] Isinkaye, F. O., Folajimi, Y. O., dan Ojokoh, B. A. (2015). *Recommendation Systems: Principles, Methods and Evaluation*. Egyptian Informatics Journal, 16(3), 261-273.
- [2] Junarto. (2015). *Perancangan dan Pembuatan Aplikasi Data Mining Berbasis Web Menggunakan Algoritma C.45 untuk Memprediksi Kelulusan Mahasiswa S-1 Pada Universitas Darma Persada*. Skripsi, Fakultas Teknik, Universitas Darma Persada.
- [3] Rissino, S., dan Lambert-Torres, G. (2009). *Rough Set Theory—Fundamental Concepts, Principals, Data Extraction, and Applications*. Data Mining and Knowledge Discovery in Real Life Applications. ISBN 978-90263-53-0. pp. 438, I-Tech.
- [4] Komorowski, J., Pawlak, Z., Polkowski, L. dan Skowron, A. (2002). *Rough Sets: A Tutorial*. URL: alfa.mimuw.edu.pl/prace/1999/D5/Tutor06 09.ps.
- [5] Agrawal, R., dan Srikant, R. (1994). *Fast Algorithms for Mining Association Rules*. Proc. 20th int. conf. Very Large Data Bases, VLDB. Vol. 1215, pp. 487-499.
- [6] Gunadi, G., dan Sensuse, D.I. (2012). *Penerapan Metode Data Mining Market Basket Analisis Terhadap Data Penjualan Produk Buku dengan Menggunakan Algoritma Apriori dan Frequent Pattern Growth (FP-Growth)*. Jurnal Telematika MKOM. Vol. 4, No. 1, Hal 118-132.
- [7] Zhao, Y. (2011). R and data mining: Examples and case studies. *R and Data Mining*, 1-4.