

**PENERAPAN ALGORITMA APRIORI UNTUK MENEMUKAN HUBUNGAN ANTARA
FAKTOR CUACA DAN CURAH HUJAN
(STUDI KASUS: KOTA TANJUNGPINANG)**

Seravina Sandra Purwandari¹, Martaleli Bettiza², Alena Uperiati³
seravinasp@gmail.com

Program studi Teknik Informatika, Fakultas Teknik, Universitas Maritim Raja Ali Haji

Abstract

Observation of weather factors is carried out by the Meteorology, Climatology and Geophysics Agency (BMKG) using sophisticated tools to then use the data for analysis and forecasting. Data from observations of weather factors are increasing day by day, which can cause data accumulation. One of the data mining algorithms that can be used in this case is the a priori algorithm. This algorithm is used to form an association rule and the value of interestingness measurement to determine the correlation between weather attributes and rainfall. Based on the calculation, it is known that with a minimum support value of 0.4 and a minimum confidence 0.8, an association rule of 3 itemset is obtained with values of lift, conviction and leverage that describe positive correlation between items, namely if the air humidity is moderate (80-89), the wind speed is weak (1.6-3.3) then the rainfall is light and if the temperature is moderate (25.4-27.8), moderate air humidity (80-89) then rainfall is light.

Keywords: algoritma apriori, interestingness measurement, weather factors, rainfall

I. Pendahuluan

Pengamatan terhadap faktor-faktor cuaca dilakukan oleh Badan Meteorologi Klimatologi dan Geofisika (BMKG). Pengamatan tersebut dilakukan secara otomatis menggunakan alat-alat canggih untuk mendukung data analisis dan prakiraan. Beberapa faktor cuaca yang diamati yaitu suhu, kelembapan udara, kecepatan angin, arah angin dan curah hujan. Data faktor cuaca yang dikumpulkan tersebut kemudian diolah sehingga menghasilkan prakiraan yang dibutuhkan. Data hasil pengamatan faktor cuaca semakin hari semakin bertambah sehingga dapat menyebabkan penumpukan data. Sedangkan untuk membuat prakiraan dibutuhkan data pengamatan tersebut. Salah satu teknik yang dapat diterapkan yaitu association rule. Association rule adalah salah satu metode data mining yang dapat mengidentifikasi hubungan antar item. Salah satu algoritma yang biasa dipakai adalah algoritma apriori. (Fauzy et al, 2016)

Ada banyak penelitian terdahulu yang membahas tentang prakiraan curah hujan, termasuk bidang data mining. Penelitian dengan metode association rule sudah pernah dilakukan sebelumnya oleh Khasanah dan Nhita (2018) yang melakukan penelitian tentang prakiraan hujan dengan menggunakan metode association rule namun menerapkan algoritma FP-Growth. Penelitian lain dilakukan oleh Delfrianti, dkk (2020). Penelitian ini menggunakan metode yang sama yaitu association rule. Meskipun objek penelitiannya berbeda namun penelitian tersebut juga menggunakan nilai interestingness measurement pada association rule-nya. Penelitian ini juga menerapkan nilai support, confidence, lift, conviction dan leverage. Pada penelitian ini, penggunaan algoritma apriori bertujuan untuk menemukan association rule. Nilai minimum support dan

confidence ditujukan untuk membentuk kombinasi itemset dan nilai interestingness measurement lift, conviction dan leverage dapat menunjukkan kekuatan association rule yang terbentuk. Oleh karena itu, penulis melakukan penelitian menerapkan algoritma apriori untuk menemukan keterhubungan antara faktor cuaca dan curah hujan yang akan ditunjukkan dari item-item yang terbentuk dalam suatu association rule.

II. Metode Penelitian

2.1 Data parameter cuaca

Variabel yang akan digunakan adalah suhu rata-rata, kelembapan udara, kecepatan angin rata-rata, arah angin dan curah hujan per harinya untuk wilayah Tanjungpinang dimulai pada Januari 2017 sampai Desember 2019. Semua parameter awalnya bernilai numerik yang kemudian dikonversi ke dalam bentuk kategori. (Khasanah & Nhita, 2018).

2.2 Preprocessing

Parameter suhu dan kelembapan udara tidak memiliki kategori kelas berdasarkan dari BMKG. Oleh karena itu, data dibagi kedalam kategori kelas dengan panjang kelas yang sama menggunakan equal width. (Khasanah & Nhita, 2018)

$$equal\ width = \frac{\max\ value - \min\ value}{number\ of\ categories}$$

Parameter arah angin, kecepatan angin dan curah hujan dikonversi berdasarkan ketentuan yang ada dari pihak BMKG.

2.3 Algoritma Apriori

Algoritma apriori menggunakan pengetahuan frekuensi atribut yang telah diketahui sebelumnya untuk memproses informasi selanjutnya. Algoritma Apriori dibagi menjadi beberapa tahap proses untuk menemukan frequent itemsets oleh Han, dkk (2012).

- a. Setiap item masuk ke dalam anggota dari kandidat itemset berukuran 1 (C1). Pada proses ini, juga dilakukan perhitungan *support* setiap item.
- b. Menentukan item dalam C1 yang memenuhi minimum *support*, yang selanjutnya disebut dengan itemset frekuensi berukuran 1 (L1).
- c. Algoritma Apriori menggunakan penggabungan L1 dan L1 ($L1 \bowtie L1$) untuk membangkitkan kandidat itemset berukuran 2 (C2). Pada tahap ini, juga dilakukan perhitungan *support* untuk setiap kandidat itemset berukuran 2.
- d. Itemset frekuensi berukuran 2 (L2) ditentukan dengan memilih kandidat itemset berukuran 2 (C2) yang memenuhi minimum *support*.
- e. Menggabungkan L2 dan L2 ($L2 \bowtie L2$) untuk membangkitkan kandidat itemset berukuran 3 (C3) dan menentukan nilai *support* masing-masing kandidat itemset berukuran 3 (C3).
- f. Menentukan itemset frekuensi berukuran 3 (L3) dengan memilih kandidat itemset berukuran 3 (C3) yang memenuhi minimum *support*.
- g. Akhir dari proses ini adalah ditemukannya seluruh *frequent* itemsets.

2.3 Interestingness measurement

Association rule yang terbentuk perlu diuji kekuatannya dengan interestingness measure. Interestingness measure merupakan hal penting dalam data mining. Pengukuran ketertarikan atau interestingness measure yang digunakan dalam penelitian ini adalah :

a. Support

Support adalah nilai penunjang atau persentase kombinasi item dalam database. Support sering digunakan untuk menghilangkan aturan yang tidak menarik.

$$Support(A) = \frac{t(A)}{t} \quad (1)$$

Sedangkan nilai support untuk 2 item adalah :

$$\text{Support}(A \rightarrow B) = \text{Support}(A \cup B) = \frac{t(A \cup B)}{t} \quad (2)$$

Keterangan :

$t(A)$ = jumlah transaksi mengandung *item* A

$t(A \cup B)$ = jumlah transaksi yang mengandung *item* A dan B

t = jumlah total transaksi

b. Confidence

Confidence adalah nilai kepastian yaitu kuatnya hubungan antara-item dalam association rule.

$$\text{Confidence}(A \rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A)} \quad (3)$$

c. Lift

Lift adalah nilai yang mengukur besarnya hubungan antara antecedent dan consequent yang tidak saling bergantung (independent).

$$\text{Lift}(A \rightarrow B) = \frac{\text{Support}(A \cup B)}{\text{Support}(A) * \text{Support}(B)} \quad (4)$$

d. Conviction

Conviction sangat memperhatikan arah dari suatu association rule. Conviction mengindikasikan bahwa $\text{conviction}(A \rightarrow B) \neq \text{conviction}(B \rightarrow A)$.

$$\text{conv}(A \rightarrow B) = \frac{1 - \text{Support}(B)}{1 - \text{conf}(A \rightarrow B)} \quad (5)$$

e. Leverage

Leverage adalah nilai yang mengukur seberapa banyak item antecedent dan consequent yang dijual bersamaan dari suatu data set. Nilai positif menandakan hubungan yang positif antara item begitupun sebaliknya.

$$\text{leve}(A \rightarrow B) = \text{Support}(A \rightarrow B) - \text{Support}(A) \times \text{Support}(B) \quad (6)$$

Kelima interestingness measurement tersebut yang akan digunakan dalam penelitian ini. Berdasarkan nilai dari parameter diatas dapat dilihat hubungan atau korelasi antar item pada suatu association rule.

2.4 Prosedur Penelitian

Adapun Prosedur Penelitian yang dilakukan sebagai berikut:

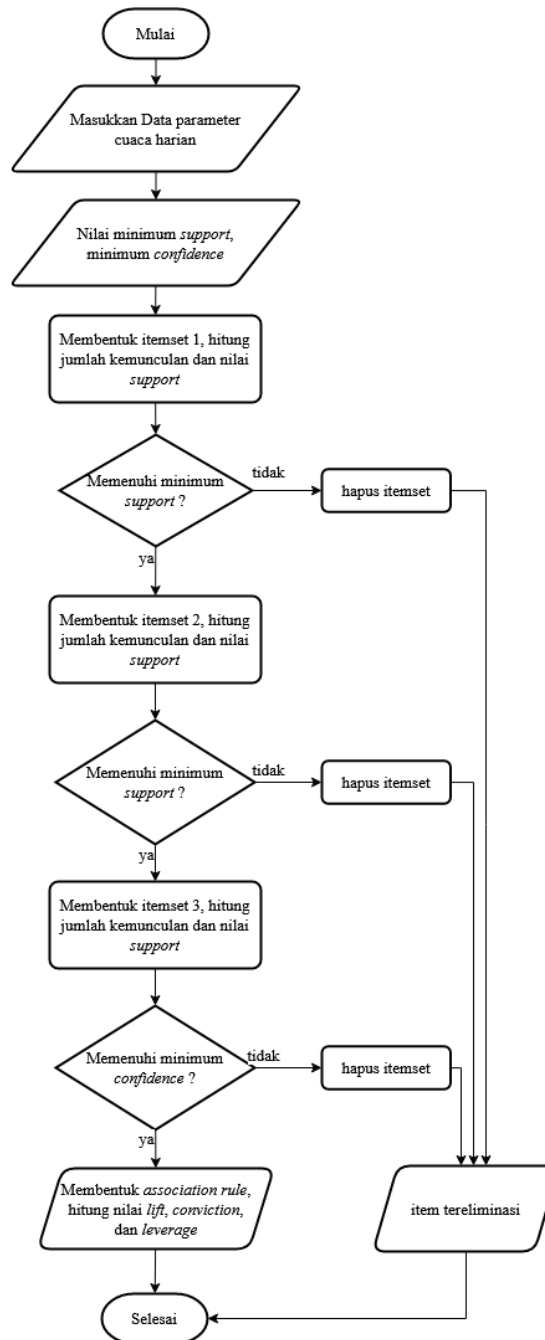


Gambar 1. Prosedur Penelitian

Pada Prosedur Penelitian Gambar 1, menunjukkan urutan tahapan kegiatan yang dilakukan dalam penelitian ini. Tahapan dimulai dari kegiatan studi pustaka hingga memperoleh hasil penelitian. Setelah melalui tahap analisa dan implementasi sistem akan mulai diuji outputnya. Tahapan tersebut dilakukan guna memperoleh hasil yang maksimal.

2.5 Flowchart Sistem

Adapun Flowchart Sistem yang dilakukan seperti terlihat pada Gambar 2.



Gambar 2. Flowchart Sistem

Pada proses perancangan dan analisis penelitian ini digambarkan pada Gambar 2 sebagai gambaran awal proses perancangan sistem. Data yang dimasukkan akan dihitung nilai support masing-masing item untuk kemudian dieliminasi apabila tidak memenuhi nilai minimum support. Setelah dilakukan hingga terbentuk kombinasi 3 itemset, kemudian akan dibentuk association rule berdasarkan 3 itemset tersebut. Association rule yang terbentuk adalah yang memenuhi nilai minimum confidence.

III. Hasil dan Pembahasan

3.1 Hasil Penelitian

Data yang digunakan adalah data balita sebanyak 1095 data. Data yang digunakan adalah parameter cuaca yaitu curah hujan, suhu, kelembapan udara, arah angin dan kecepatan angin per harinya.

3.2 Algoritma Apriori

Proses pengujian akan dilakukan dengan menggunakan nilai minimum support adalah 0.4 dan nilai minimum confidence adalah 0.8. Perhitungan nilai support dilakukan dengan persamaan (1). Contoh berikut menghitung nilai support item r.ringan.

$$Support(r.ringan) = \frac{943}{1095} = 0.86$$

Hasil perhitungan itemset 1 yang diperoleh adalah sebagai berikut:

Tabel 1. Itemset 1

No	Item	Jumlah kemunculan	Support
1	r.ringan	943	0.86
2	t.sedang	767	0.70
3	h.sedang	694	0.63
4	s.sepoi lemah	689	0.63

Tabel 1 menunjukkan itemset 1 yang memenuhi nilai minimum support. Setelah itemset 1 terbentuk, perhitungan dilanjutkan untuk menghasilkan pembentukan itemset 2. Nilai support itemset 2 dihitung dengan menggunakan persamaan (2) seperti contoh berikut:

$$Support(r.ringan, t.sedang) = \frac{662}{1095} = 0.60$$

Hasil perhitungan itemset 2 yang diperoleh adalah sebagai berikut:

Tabel 2. Itemset 2

No	Item	Jumlah kemunculan	Support
1	r.ringan, t.sedang	662	0.60
2	r.ringan, h.sedang	621	0.57
3	r.ringan, s.sepoi lemah	612	0.56
4	t.sedang, h.sedang	515	0.47
5	t.sedang, s.sepoi lemah	449	0.41
6	h.sedang, s.sepoi lemah	484	0.44

Tabel 2 menunjukkan kombinasi itemset 2 yang memenuhi nilai support. Itemset 2 yang memenuhi kondisi nilai support minimum dilanjutkan untuk diproses membentuk itemset 3. Nilai support itemset 3 dihitung dengan menggunakan persamaan (2).

Tabel 3. Itemset 3

No	Item	Jumlah kemunculan	Support
1	r.ringan, t.sedang, h.sedang	457	0.42
2	r.ringan, h.sedang, s.sepoi lemah	439	0.40

Itemset 3 yang memenuhi nilai minimum support kemudian akan dibentuk menjadi association rule jika memenuhi nilai minimum confidence. Association rule terbentuk akan disertai dengan nilai lift, conviction dan leverage. Nilai confidence dihitung menggunakan persamaan (3), lift menggunakan persamaan (4), conviction menggunakan persamaan (5) dan leverage menggunakan persamaan (6).

Tabel 4. Hasil association rule

No	Itemset	support (x => y)	Conf. (x => y)	Lift (x => y)	conviction (x => y)	leverage (x => y)
1	h.sedang, s.sepoi lemah =>	0.4	0.91	1.05	6.44	0.02

	r.ringan					
2	t.sedang , h.sedang => r.ringan	0.42	0.89	1.03	5.17	0.01

3.3 Pembahasan

Association rule terbentuk berdasarkan nilai minimum confidence. Rule yang terbentuk disertai nilai lift, conviction dan leverage. Nilai lift, conviction dan leverage digunakan untuk melihat keterkaitan antar item yang dimiliki rule tersebut. Keterikatan antar item juga digambarkan dari nilai confidence rule.

Nilai confidence menggambarkan nilai kepercayaan association rule. Semakin tinggi nilai confidence maka semakin tinggi kekuatan association rule tersebut. Nilai confidence berada pada range 0 sampai 1. Oleh karena itu, association rule yang terbentuk pada tabel diatas memiliki kekuatan rule yang baik karena memiliki nilai confidence 0.91 dan 0.89. Nilai confidence menunjukkan besar kemunculan X dimana Y juga muncul.

Lift adalah ukuran korelasi untuk menentukan tingkat kekuatan association rule. Semakin tinggi nilai lift maka rule tersebut semakin berpengaruh. Nilai lift lebih dari 1 menunjukkan bahwa association rule tersebut lebih sering muncul bersama. Lift memiliki range dari 0 sampai ∞ . Oleh karena itu association rule pada tabel diatas dapat dikatakan valid karena keeratn antar item yang kuat dan positif dibuktikan dengan nilai lift yang dihasilkan lebih dari 1 yaitu 1.05 dan 1.03.

Conviction dapat menggambarkan kekuatan dan akurasi minimum rule. Semakin tinggi nilai conviction maka akan semakin baik dan akurat rule tersebut. Association rule yang memiliki nilai conviction semakin menjauh dari 1 bahkan sampai tak hingga maka akan dianggap semakin akurat. Nilai conviction 1 menunjukkan sifat independen. Maka nilai conviction yang diperoleh rule adalah 6.44 dan 5.17 menunjukkan keterhubungan atau dependen rule tersebut.

Leverage menggambarkan banyaknya X dan Y muncul secara bersamaan dalam suatu database. Nilai positif menggambarkan hubungan yang positif begitupun sebaliknya. Nilai leverage yang diperoleh adalah 0.02 dan 0.01. Oleh karena itu nilai positif tersebut menggambarkan keterikatan antar item pada rule tersebut.

IV. Kesimpulan

Kesimpulan yang dapat diambil pada penelitian ini adalah :

1. Penerapan metode association rule terhadap faktor cuaca menghasilkan association rule dengan 3 itemset yang berkolerasi positif yaitu jika kelembapan udara sedang (80-89), kecepatan angin sepoi lemah (1.6-3.3) maka curah hujan ringan dan jika suhu sedang (25.4-27.8), kelembapan udara sedang (80-89) maka curah hujan ringan.
2. Penerapan algoritma apriori dan *interestingness measures* dapat menunjukkan bahwa *association rule* yang dihasilkan saling memiliki keterhubungan yang positif antara faktor cuaca dengan curah hujan.

V. Daftar Pustaka

- Anhar. 2010. PHP dan MySQL secara otodidak. PT. Transmedia: Jakarta.
- Delfitriani D, Rahmaningrum V, dan Ginantaka A. 2020. Desain Afektif Kemasan Produk Household Care Berbahan Baku Serat Selulosa pada PT XY. *Industria: Jurnal Teknologi dan Manajemen Agroindustri*, 9(2).
- Fauzy M, Asror I. 2016. Penerapan metode association rule menggunakan algoritma Apriori pada simulasi prakiraan hujan wilayah kota Bandung. *Jurnal Ilmiah Teknologi Infomasi Terapan*, 2(3).

- Han J, Kamber M. 2006. *Data Mining : Concept and Techniques*, 2nd edition. Morgan Kauffman: San Fransisco.
- Han J, Kamber M, dan Pei Jian. 2012. *Data Mining : Concepts and Techniques*. Elsevier Inc.
- Hermanto LA. 2018. Prakiraan Tinggi Gelombang Air Laut Menggunakan Data Mining. *Jurnal IPTEK*, Hal. 37-44.
- Khasanah FN, Nhita F. 2018. Weather Forecasting in Bandung Regency based on FP-Growth Algorithm. *International Journal on Information and Communication Technology (IJoICT)*, 4(2), 1-10.
- Kusrini, Luthfi ET. 2009. *Algoritma Data Mining*. Andi Offset: Yogyakarta.
- Larose, TD. 2005. *Discovering Knowledge in Data: An Introduction to Data Mining*. John Willey & Sons Inc.
- Listriani D, Setyaningrum AH, dan A FEM. 2016. Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Analisa Pola Belanja Konsumen. *Jurnal Teknik Informatika*, Vol.9, No.2.
- Manimaran J, Velmurugan T. 2015. Analysing the quality of Association rules by Computing an Interestingness measures. *Indian Journal of Science and Technology*, Vol. 8.
- Pratama HC, Bettiza M, dan Matulatan T. 2014. Penerapan Algoritma Apriori Dalam Menemukan Hubungan Data Awal Masuk Mahasiswa Dengan Prestasi Akademik Studi Kasus: STAI Miftahul Ulum Tanjungpinang. Jurusan Teknik Informatika, Fakultas Teknik, Universitas Maritim Raja Ali Haji: Tanjungpinang.
- Putri LD. 2016. Rekomendasi Aktivitas Pemasaran Berdasarkan Penerapan Association Rules Pada Minimarket X. Doctoral dissertation. Institut Teknologi Sepuluh Nopember.
- Sari P, Sinaga B. 2018. Aplikasi data mining dengan menggunakan algoritma apriori untuk penjualan produk terbesar pada Cv. Sakura Photo. *Jurnal Mantik Penusa*, Vol. 22.
- Setiawan DD. 2009. Penggunaan Metode Apriori Untuk Analisa Keranjang Pasar Pada Data Transaksi Penjualan Minimarket Menggunakan Java & MySQL. Skripsi Program Studi Teknik Informatika, Universitas Gunadarma.
- Vishwakarma G, Partey S, Joshi H, Patel J, dan Jangid P. 2019. Online Sequential Behaviour Analysis using Apriori Algorithm. *International Research Journal Of Engineering And Technology(IRJET)*, Vol. 06.
- Zulaikha S, Bettiza M, dan Ritha N. 2020. Applying Improved Apriori Algorithm in Figuring out the Relation between Weather Factors and Rainfall. *Journal of Innovation and Technology*, 1(1), 21-25.

VI. Ucapan Terimakasih

Penulis mengucapkan terima kasih yang sebesar-besarnya dan penghargaan yang mulia kepada :

1. Tuhan Yang Maha Esa yang telah mendengar serta mengabulkan doa dan memberikan petunjuk hingga kemudahan.
2. Kedua orang tua yang senantiasa memberikan doa dan dukungan demi kelancaran penyelesaian tugas akhir ini.
3. Adik-adik serta seluruh keluarga dan kerabat yang telah mendukung dan mendoakan.
4. Ibu Nurul Hayaty, S.T.,M.Cs. selaku ketua jurusan program studi Teknik Informatika Universitas Maritim Raja Ali Haji.
5. Ibu Martaleli Bettiza, S.Si., M.Sc., selaku dosen Pembimbing Skripsi I yang telah membimbing dan mengarahkan penulis selama penyusunan tugas akhir.
6. Ibu Alena Uperiati, S.T.,M.Cs., selaku dosen Pembimbing Skripsi II yang juga selalu memberikan bimbingan dan arahan kepada penulis selama penyusunan tugas akhir.
7. Seluruh dosen Fakultas Teknik yang telah memberikan ilmu yang tak terbatas selama perkuliahan.
8. Ibu Vivi selaku perwakilan dari pihak BMKG Kota Tanjungpinang yang sudah membantu.

9. Rekan-rekan seperjuangan Teknik Informatika 2016 atas jalinan persahabatan dan informasi serta bantuan yang diberikan kepada penulis.
10. Serta semua pihak yang tidak dapat disebutkan satu persatu yang telah membantu baik secara langsung maupun tidak langsung.