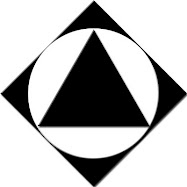
**Data Mining dan Retrieval**

**Tugas 4**

**ASOSIASI**

****

**Disusun oleh**

**Kelompok 6 :**

**Jeffry Sukmawidjaja - 152022009   
Katon Rinantomo - 152022012   
Raihan Alrais Rahman - 152022035   
Mohammad Rohman - 152022039**

**Prodi Informatika**

**Fakultas Teknologi Industri**

**Institut Teknologi Nasional Bandung**

**BAB 1**

**PENDAHULUAN**

**1.1 Latar Belakang**

Dalam data mining, model asosiasi adalah metode analisis yang digunakan untuk menemukan pola atau hubungan antar item dalam kumpulan data yang sangat besar. Tujuan dari metode ini adalah untuk menemukan item yang sering muncul bersamaan dalam set data.

Salah satu algoritma yang paling populer dalam model asosiasi adalah algoritma Apriori, yang bekerja dengan mencari pola itemset yang sering muncul atau itemset yang sering muncul.

**1.2 Rumusan Masalah**

**●** Bagaimana definisi dan sejarah dari model asosiasi?

**●** Apa saja formula yang umum digunakan dalam model asosiasi?

**●** Bagaimana contoh penerapan model asosiasi menggunakan Excel?

**●** Bagaimana implementasi program asosiasi sederhana menggunakan Python?

**1.3 Tujuan Penulisan**

**●** Mengidentifikasi konsep dasar model asosiasi .

**●** Menjelaskan berbagai rumus yang digunakan.

**●** Menyajikan contoh penerapan di Excel.

**●** Mengimplementasikan model asosiasi sederhana dengan Python.

**BAB 2**

**DEFINISI DAN SEJARAH APRIORI**

**2.1 Definisi Apriori**

Algoritma Apriori sangat populer dalam mining aturan asosiasi, yaitu proses menemukan aturan atau asosiasi antar item dalam kumpulan data transaksi. Metode data mining ini digunakan untuk menemukan pola atau hubungan antar-item dalam kumpulan data yang sangat besar.

Prinsip dasar dari algoritma Apriori adalah konsep *apriori property*, yang menyatakan bahwa jika sebuah itemset (kumpulan item) sering muncul dalam data, maka semua subset dari itemset tersebut juga harus sering muncul.

Salah satu contoh aplikasinya adalah analisis pasar keranjang, di mana algoritma Apriori membantu mengidentifikasi produk yang sering dibeli oleh konsumen. Misalnya, sebuah swalayan memiliki database tentang barang yang dibeli pelanggan di toko mereka. Dengan menggunakan algoritma apriori, swalayan berusaha mengungkap hubungan dan pola dari barang-barang yang dijual. Hasilnya, swalayan tersebut mengetahui bahwa ketika pelanggan membeli kopi, mereka juga membeli gula.

Berikut ini adalah langkah-langkah dari algoritma Apriori:

1. Hitung support dari itemset (dengan ukuran k = 1) dalam database. Proses ini akan menghasilkan himpunan kandidat.
2. Pangkas (pruning) kumpulan kandidat dengan cara menghilangkan item yang memiliki support lebih kecil dari ambang batas (threshold) yang diberikan.
3. Gabungkan itemset yang paling sering muncul untuk membentuk himpunan berukuran k + 1, dan ulangi himpunan di atas sampai tidak ada lagi itemset yang dapat dibentuk.

**2.2 Sejarah Apriori**

* 1993 - 1994

**Rakesh Agrawal dan Ramakrishnan Srikant** dari IBM Almaden Research Center memperkenalkan **algoritma Apriori**. Karya mereka ditujukan untuk membuat association rule mining menjadi lebih efisien untuk database transaksional, terutama untuk aplikasi ritel seperti analisis keranjang pasar. Algoritma ini memperkenalkan teknik “pemangkasan” berdasarkan “properti penutupan ke bawah” (frequent itemsets harus memiliki subset yang sering) untuk menghilangkan perhitungan yang tidak perlu.

* 1996 - 1997

Makalah dan implementasi lanjutan mengidentifikasi keterbatasan Apriori dalam hal kinerja pada dataset besar. Hasilnya, **AprioriTid** dan **Apriori Hybrid** dikembangkan, yang mengoptimalkan algoritma dengan mengganti teknik pemindaian data yang berbeda tergantung pada ukuran dataset dan frekuensi itemset. Hal ini semakin meningkatkan efisiensi dan membuat Apriori dapat diterapkan pada basis data yang lebih besar.

* 2000 - Sekarang

Komunitas data mining mengakui Apriori sebagai algoritma dasar untuk pembelajaran aturan asosiasi. Pendekatannya mengilhami banyak variasi, termasuk algoritma yang lebih efisien seperti FP-Growth. FP-Growth mengatasi masalah kinerja Apriori dengan menghindari pemindaian basis data yang berulang-ulang, dengan menggunakan struktur data terkompresi (FP-tree) sebagai gantinya.

**BAB 3**

**FORMULA A PRIORI**

**3.1 Mencari Support Item Pada Item Set**

Support adalah rasio jumlah transaksi yang mengandung item set tertentu terhadap total jumlah transaksi dalam dataset. Secara matematis, support untuk item set X didefinisikan sebagai:  


**3.2 Mencari Confidence**

Confidence memberikan gambaran dari setiap rule yang dibuat. Pada saat confidence kedua rule dikatakan tinggi maka ada kemungkinan pembeli untuk membeli 2 barang yang sama pada saat bersamaan. Berikut merupakan rumus untuk mencari Confidence:



**BAB 4**

**STUDI KASUS**

**4.1 Perhitungan Manual**

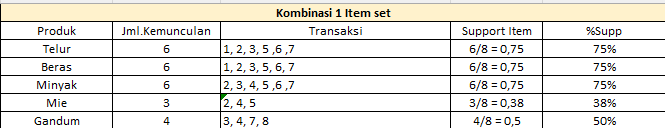
Didapatkan data dengan 10 transaksi yang berbeda - beda jenis nya di setiap interaksi. Terdapat 5 item yaitu : Telur, Beras, Minyak, Mie dan Gandum.

****

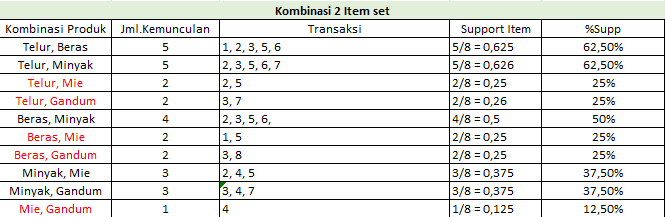
Setelah itu kita buatkan data tabular untuk melihat record di setiap transaksi.

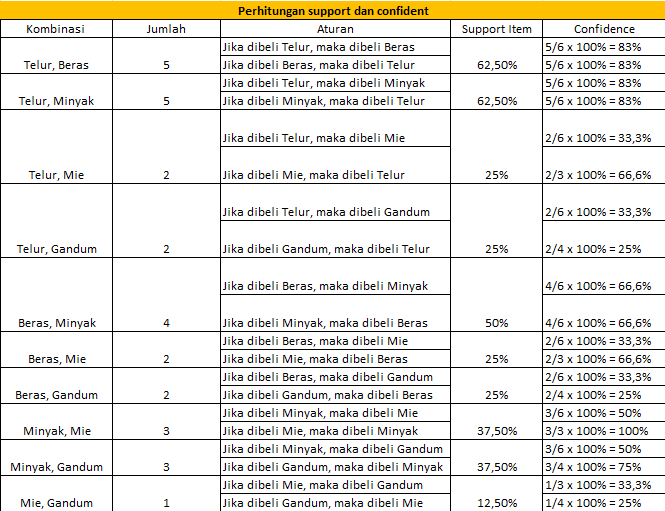
****

| Lalu kita lakukan asosiasi dengan 1 item set yaitu telur, beras, minyak, mie dan gandum. Lalu hitung support itemnya untuk setiap item dan bandingkan hasil persentase nya sesuai dengan threshold yang ditentukan. Aturan asosiasi pada tabel transaksi minimal support itemnya adalah 0,3 (30%) dan nilai = 3. | | |
| --- | --- | --- |
|
|

****

Setelah Melakukan threshold, lanjut melakukan asosiasi dengan 2 item set. Sesuai dengan judulnya 2 item set, maka kita akan memilih kemungkinan yang akan dipakai untuk setiap data.

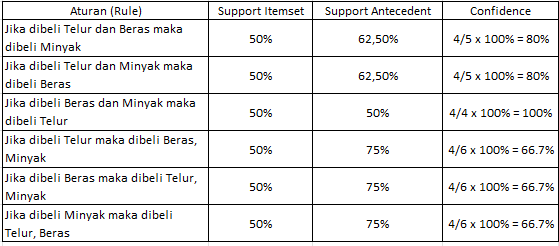


Terlihat setelah diasosiasikan, hasil support terdapat lebih kecil dari 30%. Setelah itu kita cari nilai confidence berdasarkan aturan yang ditentukan untuk setiap A dan B. Penulis menetapkan nilai threshold untuk Confidence pada studi kasus ini adalah 60%. Maka untuk setiap pembelian dengan nilai confidence diatas 60% akan memberikan kepastian bahwa pembeli akan membeli 2 item yang berkaitan secara bersamaan

Setelah itu kita lanjut dengan kombinasi 3 item set dimana 3 item set ini di tentukan dari nilai support item pada langkah 2 item set yang hasil support itemnya diatas 30%.



Karena hasil asosiasi berdasarkan threshold hanya Telur, Beras, dan Minyak yang diatas 30%, maka aturan (rule) pada kombinasi item set hanya dari kombinasi Telur, Beras, Minyak. karena nilai confidence nya diatas 60% maka ada kemungkinan pembeli akan membeli 3 item berkaitan secara bersamaan.



**4.2 Implementasi Program**

**1. import pandas as pd**

**2. from apyori import apriori**

**3.**

**4. data = pd.read\_csv('transaksi2.csv')**

**5.**

**6. data['Items'] = data['Items'].apply(lambda x: x.split(','))**

**7. transactions = data['Items'].tolist()**

**8.**

**9. min\_support = 0.3**

**10. min\_confidence = 0.6**

**11.**

**12. results = list(apriori(transactions, min\_support=min\_support, min\_confidence=min\_confidence))**

**13.**

**14. for item in results:**

**15. pair = item.items**

**16. items = [x for x in pair]**

**17. if len(items) == 3:**

**18. print(f"Frequent Itemset: {items}")**

**19. print(f"Support: {item.support}")**

**20.**

**21. for rule in item.ordered\_statistics:**

**22. if rule.confidence >= min\_confidence:**

**23. antecedent = [x for x in rule.items\_base]**

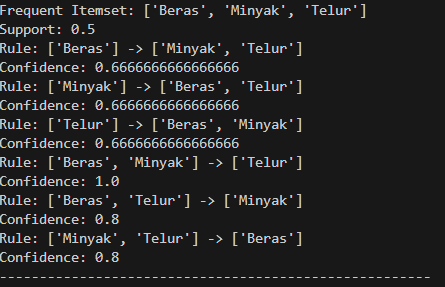
**24. consequent = [x for x in rule.items\_add]**

**25. print(f"Rule: {antecedent} -> {consequent}")**

**26. print(f"Confidence: {rule.confidence}")**

**27. print("------------------------------------------------------")**

Output dari program:

****