

# Analisis Data Mining Produk Retail Menggunakan Metode Asosiasi Dengan Menerapkan Algoritma Apriori

Oleh:

Mochammad Septa Sandy,  
Hamzah Setiawan, S.Kom., M.Kom.

Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

April, 2023



# Pendahuluan

## A. Latar Belakang Masalah

Pada bisnis produk retail ini, yang sering ditemui adalah masalah satu barang yang selalu menumpuk karena jumlah pembelian yang saat ini sedang menurun[3]. Oleh karena itu, perusahaan dapat melakukan analisis untuk mengetahui apa yang terkait dengan data yang dimilikinya sebanyak 541.910 pembelian produk retail oleh konsumen yang berupa dataset yang saya ambil data melalui online dari github yang berupa csv menggunakan jypter notebook. Dengan hasil tersebut, penelitian yang dapat diperoleh harapannya suatu saat dapat membantu untuk mengatur strategi pada tata letak produk agar saling berdekatan, sehingga dapat mempengaruhi pertumbuhan prospek pembelian dari konsumen yang semakin tinggi.

# Pendahuluan

Salah satu cara produk retail ini dapat diketahui informasi pada pola belanja pelanggan adalah untuk mengetahui kebiasaan pelanggan dalam membeli barang yang dilakukan secara bersama-sama. Setiap pelanggan memiliki kebiasaan membeli barang pada waktu yang sama. Informasi tersebut bisa mengatasi permasalahan yang sering dihadapi oleh perusahaan yaitu keluhan pada pelanggan ketika terjadi persediaan kosong yang sering sekali dibeli secara bersama-sama[2].

Dengan menggunakan teknik data mining ini masih banyak juga cara yang dapat digunakan untuk mengatasi masalah tersebut. Salah satu metode yang dapat digunakan untuk memecahkan masalah dalam data mining adalah algoritma apriori. Algoritma Apriori adalah algoritma yang menjadi bagian dari asosiasi data mining. Proses penentuan hubungan asosiasi untuk setiap item didasarkan pada nilai support dan nilai confidence yang telah ditentukan sebelumnya[7].



# Metode Penelitian

- Tahapan Penelitian

Pada dasarnya penelitian ini peneliti menggunakan aplikasi Anaconda Navigator 3 2022 dengan IDLE Jupyter Notebook versi 6.4.8 untuk melakukan pemrosesan pada dataset ini dan dilakukan pengambilan dataset melalui Github untuk data penjualan produk retail tersebut. Untuk mempermudah proses Data Mining, dari tabel data transaksi bulan Desember 2010 sampai dengan Desember 2011 yang menjadi objek penelitian ini.

Tool yang memungkinkan seseorang untuk melakukan analisis dan mengolah data disebut jupyter notebook. Jupyter notebook dapat digunakan untuk menulis kode, menjalankan kode yang dituliskan dan melihat hasil dari kode yang dijalankan dalam satu tampilan yang interaktif. Output dari kode yang diinput bisa berupa teks sederhana ataupun dalam bentuk grafik.[8]

# Metode Penelitian

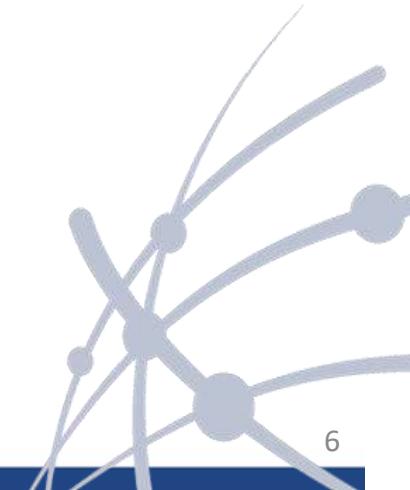
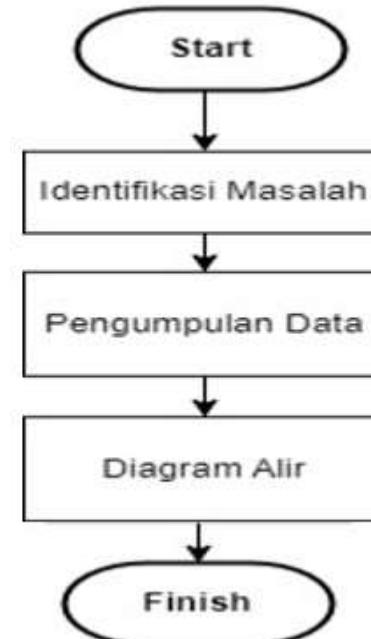
Algoritma apriori direkomendasikan untuk meningkatkan penjualan saat ini dengan cara asosiasi otomatis, yang dapat direkomendasikan ke produk saat transaksi dilakukan[5]. Algoritma Apriori ini diharapkan dapat memberikan dukungan pada keputusan antar barang yang dibeli secara bersama oleh pelanggan. Data Mining adalah proses yang memerintahkan satu atau lebih pembelajaran (komputer)[6]

Data mining adalah aktivitas yang mencakup pengumpulan, penggunaan data penjualan untuk menemukan keteraturan, pola, dan hubungan dalam kumpulan data yang besar[1]. Data mining digunakan untuk mengekstraksi informasi penting yang tersembunyi di perusahaan. Dengan adanya data mining akan mendapatkan kumpulan data berupa pengetahuan dalam jumlah kumpulan data yang banyak[9].

# Hasil dan Pembahasan

- Tahapan Dataset

Pada flowchart dibawah ini dapat dilihat ada 5 proses yaitu proses penelitian yang dilakukan menggunakan tahapan pengumpulan data, membersihkan data, memisahkan data sesuai wilayah, pengkodean data, dan analisis hasil yang berada ditahapan diagram alir



# Hasil dan Pembahasan

- Tahap Pengumpulan Data

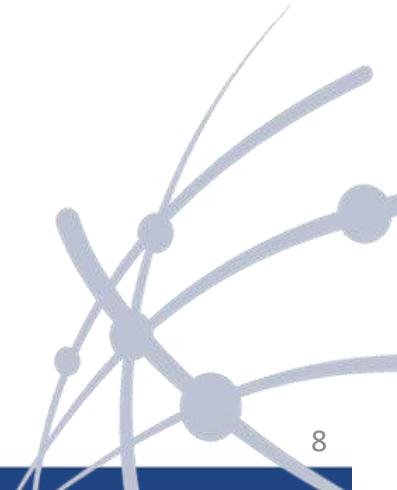
Dalam Pengumpulan Data yang berada di tahapan diatas adalah untuk melakukan pengekploran data yang ada di dalam data yang dimilikinya sebanyak 541.910 pembelian produk retail oleh konsumen yang berupa dataset yang saya ambil data melalui online dari github yang berupa csv menggunakan jypter notebook tersebut di komputasi 5 data yang paling atas.

	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

# Diagram Alir Penelitian

- Membersihkan Data

Dalam proses membersihkan data yang sudah di cleaning biar hasil lebih optimal berada di tahapan diatas adalah untuk melakukan penghapusan data yang kurang jelas semisal banyak spasi itu kita hilangkan dengan dilakukannya drooping pada data produk retail tersebut.



# Diagram Alir Penelitian

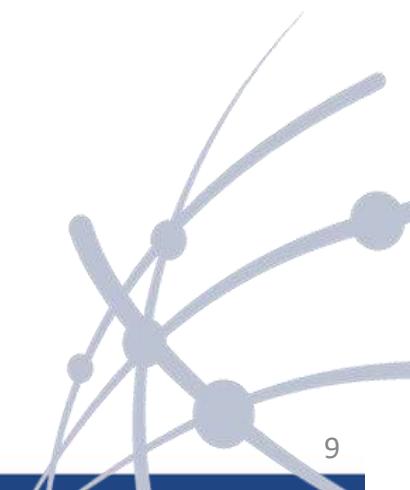
- Memisahkan Data

Dalam proses memisahkan data yang berada di tahapan diatas adalah untuk melakukan pemisahan data sesuai yang diambil dari negara mana untuk transaksi seperti wilayah transaksi tersebut yang disesuaikan dengan grup masing masing nanti kita ambil 4 sample semisal kita print dari salah satu negara data dari United Kingdom.

Out[16]:

Description	*Boombbox Ipod Classic	*USB Office Mirror Ball	10 COLOUR SPACEBOY PEN	12 COLOURED PARTY BALLOONS	12 DAISY PEGS IN WOOD BOX	12 EGG HOUSE PAINTED WOOD	HANGING EGGS HAND PAINTED	12 IVORY ROSE PEG PLACE SETTINGS	12 MESSAGE CARDS WITH ENVELOPES	12 PENCIL SMALL TUBE WOODLAND	wrongly coded 20713	wrongly coded 23343	wrongly coded- 23343	w r
InvoiceNo														
536365	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536366	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536367	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536368	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
536369	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...														
581585	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
581586	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A563185	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A563186	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
A563187	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

18667 rows × 4175 columns



# Diagram Alir Penelitian

- Pengkodean Data

Tahap pengkodean data adalah mendefinisikan fungsi pada libraries yang dilakukan pembelian bersamaan semisal tidak membeli nanti nilai menjadi 0 kalaupun ada yang membeli nanti nilai menjadi 1 berada di tahapan diatas adalah untuk mendefinisikan fungsi data agar sesuai pada suatu wilayah yang sudah di tentukan pada libraries.

```
In [ ]: def hot_encode(x):
           if(x<= 0):
           return 0
           if(x>= 1):
           return 1
```



# Analisis Hasil Penerapan Apriorinya

Pada analisis hasil Aplikasi Apriori produk retail ini untuk mengetahui informasi pola belanja pelanggan adalah mengetahui kebiasaan pelanggan dalam membeli barang yang dilakukan secara bersamaan yang diambil dari 5 aturan terkuat.

Setiap pelanggan memiliki kebiasaan membeli barang pada waktu yang bersamaan. Produk retail tidak lepas dari proses jual beli barang. Kelangsungan proses bisnis pada produk retail sangat bergantung pada jumlah penjualan barang yang dilakukan oleh toko retail yang diambil dari 4 negara terbesar untuk hasil seperti dibawah ini :



# United Kingdom

Untuk United Kingdom yang disingkat menjadi UK kita bangun untuk modelnya lalu dipanggil fungsi penerapan apriorinya dari library sebelumnya minimal supportnya 0.01 terus dibuat association rules nya digeneric literasio yang akan dilakukan pengujian modelnya setelah itu disorting berdasarkan confidence yang paling tinggi dan lift rasion paling tinggi sehingga nanti yang kita urutkan paling atas yaitu rule paling kuat kemudian ditentukan rules.head.

```
In [19]: # 1) United Kingdom
frq_items = apriori(basket_UK, min_support = 0.01, use_colnames = True)

# Mengumpulkan aturan yang disimpulkan dalam dataframe
rules = association_rules(frq_items, metric ="lift", min_threshold = 1)
rules = rules.sort_values(['confidence', 'lift'], ascending =[False, False])
print(rules.head())
```

	antecedents		consequents		\								
	117	(BEADED CRYSTAL HEART PINK ON STICK)	2019	(SUKI SHOULDER BAG, JAM MAKING SET PRINTED)	2294	(HERB MARKER MINT, HERB MARKER THYME)	2300	(HERB MARKER ROSEMARY, HERB MARKER PARSLEY)	2302	(HERB MARKER PARSLEY, HERB MARKER THYME)	(HERB MARKER ROSEMARY)	(HERB MARKER ROSEMARY)	(HERB MARKER THYME)
	antecedent	support	consequent	support	support	confidence	lift	\					
117		0.011036		0.037928	0.010768	0.975728	25.725872						
2019		0.011625		0.037928	0.011196	0.963134	25.393807						
2294		0.010714		0.012375	0.010232	0.955000	77.173095						
2300		0.011089		0.012321	0.010553	0.951691	77.240055						
2302		0.011089		0.012375	0.010553	0.951691	76.905682						
	leverage	conviction											
117	0.010349	39.637371											
2019	0.010755	26.096206											
2294	0.010099	21.947227											
2300	0.010417	20.444951											
2302	0.010416	20.443842											

# Italy

Untuk Italy yang kita bangun untuk modelnya lalu dipanggil fungsi penerapan apriorinya dari library sebelumnya minimal supportnya 0.05 terus dibuat association rules nya digeneric literasio yang akan dilakukan pengujian modelnya setelah itu disorting berdasarkan confidence yang paling tinggi dan lift rasion paling tinggi sehingga nanti yang kita urutkan paling atas yaitu rule paling kuat kemudian ditentukan rules.head.

In [20]: # 2) Italy

```
frq_items = apriori(basket_Italy, min_support = 0.05, use_colnames = True)
rules = association_rules(frq_items, metric ="lift", min_threshold = 1)
rules = rules.sort_values(['confidence', 'lift'], ascending =[False, False])
#print(rules.head())
rules.head(10)
```

Out[20]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
22	(BAKING MOULD CHOCOLATE CUPCAKES)	(RED LOVE HEART SHAPE CUP)	0.052632	0.052632	0.052632	1.0	19.0	0.049861	inf
23	(RED LOVE HEART SHAPE CUP)	(BAKING MOULD CHOCOLATE CUPCAKES)	0.052632	0.052632	0.052632	1.0	19.0	0.049861	inf
54	(BATH BUILDING BLOCK WORD)	(HOME BUILDING BLOCK WORD)	0.052632	0.052632	0.052632	1.0	19.0	0.049861	inf
55	(HOME BUILDING BLOCK WORD)	(BATH BUILDING BLOCK WORD)	0.052632	0.052632	0.052632	1.0	19.0	0.049861	inf
164	(CHRISTMAS CRAFT LITTLE FRIENDS)	(CHRISTMAS CRAFT WHITE FAIRY)	0.052632	0.052632	0.052632	1.0	19.0	0.049861	inf
165	(CHRISTMAS CRAFT WHITE FAIRY)	(CHRISTMAS CRAFT LITTLE FRIENDS)	0.052632	0.052632	0.052632	1.0	19.0	0.049861	inf
166	(CHRISTMAS GINGHAM TREE)	(CHRISTMAS GINGHAM STAR)	0.052632	0.052632	0.052632	1.0	19.0	0.049861	inf
167	(CHRISTMAS GINGHAM STAR)	(CHRISTMAS GINGHAM TREE)	0.052632	0.052632	0.052632	1.0	19.0	0.049861	inf
170	(IVORY WICKER HEART LARGE)	(CHRISTMAS GINGHAM STAR)	0.052632	0.052632	0.052632	1.0	19.0	0.049861	inf
171	(CHRISTMAS GINGHAM STAR)	(IVORY WICKER HEART LARGE)	0.052632	0.052632	0.052632	1.0	19.0	0.049861	inf

# Saudi Arabia

Untuk Saudi Arabia disingkat menjadi SA yang kita bangun untuk modelnya lalu dipanggil fungsi penerapan apriorinya dari library sebelumnya minimal supportnya 0.05 terus dibuat association rules nya digeneric literasio yang akan dilakukan pengujian modelnya setelah itu disorting berdasarkan confidence yang paling tinggi dan lift rasion paling tinggi sehingga nanti yang kita urutkan paling atas yaitu rule paling kuat kemudian ditentukan rules.head.

In [21]: # 3) Saudi Arabia

```
frq_items = apriori(basket_SA, min_support = 0.05, use_colnames = True)
rules = association_rules(frq_items, metric = "lift", min_threshold = 1)
rules = rules.sort_values(['confidence', 'lift'], ascending =[False, False])
#print(rules.head())
rules.head(6)
```

Out[21]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(ASSORTED BOTTLE TOP MAGNETS)	(GLASS JAR DAISY FRESH COTTON WOOL)	1.0	1.0	1.0	1.0	1.0	0.0	inf
1	(GLASS JAR DAISY FRESH COTTON WOOL)	(ASSORTED BOTTLE TOP MAGNETS)	1.0	1.0	1.0	1.0	1.0	0.0	inf
2	(GLASS JAR MARMALADE)	(ASSORTED BOTTLE TOP MAGNETS)	1.0	1.0	1.0	1.0	1.0	0.0	inf
3	(ASSORTED BOTTLE TOP MAGNETS)	(GLASS JAR MARMALADE)	1.0	1.0	1.0	1.0	1.0	0.0	inf
4	(GLASS JAR PEACOCK BATH SALTS)	(ASSORTED BOTTLE TOP MAGNETS)	1.0	1.0	1.0	1.0	1.0	0.0	inf
5	(ASSORTED BOTTLE TOP MAGNETS)	(GLASS JAR PEACOCK BATH SALTS)	1.0	1.0	1.0	1.0	1.0	0.0	inf

# Germany

Untuk Germany yang kita bangun untuk modelnya lalu dipanggil fungsi penerapan apriorinya dari library sebelumnya minimal supportnya 0.05 terus dibuat association rules nya digeneric literasio yang akan dilakukan pengujian modelnya setelah itu disorting berdasarkan confidence yang paling tinggi dan lift rasion paling tinggi sehingga nanti yang kita urutkan paling atas yaitu rule paling kuat kemudian ditentukan rules.head.

```
In [22]: # 4) Germany
frq_items = apriori(basket_Germany, min_support = 0.05, use_colnames = True)
rules = association_rules(frq_items, metric ="lift", min_threshold = 1)
rules = rules.sort_values(['confidence', 'lift'], ascending =[False, False])
#print(rules.head())
rules.head(3)
```

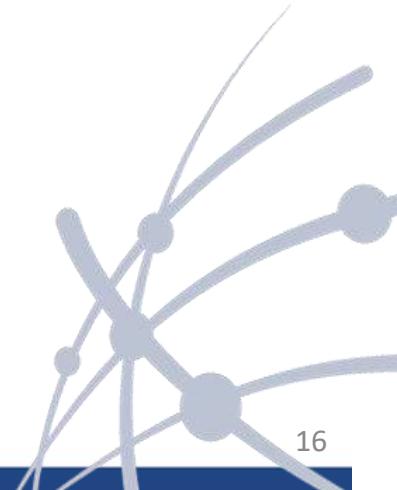
Out[22]:

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
34	(PLASTERS IN TIN STRONGMAN)	(POSTAGE)	0.070022	0.818381	0.067834	0.968750	1.183740	0.010529	5.811816
50	(RETROSPOT TEA SET CERAMIC 11 PC)	(POSTAGE)	0.056893	0.818381	0.054705	0.961538	1.174928	0.008145	4.722101
52	(ROUND SNACK BOXES SET OF 4 FRUITS)	(POSTAGE)	0.157549	0.818381	0.150985	0.958333	1.171012	0.022049	4.358862



# Kesimpulan

Penelitian ini telah berhasil melakukan seluruh tahapan penerapan apriori yang menggunakan metode asosiasi. Disini saya menggunakan 4 negara yang memilih basket united kingdom yang disingkat uk yang mempunyai minimum support 0.01, Italy yang mempunyai minimum support 0.05, terus basket Saudi Arabia yang disingkat SA yang mempunyai minimum support 0.05, dan basket germany yang mempunyai minimum support 0.05. Untuk nilai parameter yang mempunyai nilai encode 0 dan 1. Dari hasil tersebut pada penelitian ini dapat di temukan metode asosiasi yang menerapkan algoritma apriori untuk mengetahui nilai support dan hasil nilai confidence 0.968750 yang dari item tersebut.



# Referensi

- [1] A. R. Wibowo and A. Jananto, "Implementasi Data Mining Metode Asosiasi Algoritma FP-Growth Pada Perusahaan Ritel," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 2, p. 200, 2020, doi: 10.35585/inspir.v10i2.2585.
- [2] I. M. D. P. Asana, I. G. I. Sudipa, A. A. T. W. Mayun, N. P. S. Meinarni, and D. V. Waas, "Aplikasi Data Mining Asosiasi Barang Menggunakan Algoritma Apriori-TID," *INFORMAL Informatics J.*, vol. 7, no. 1, p. 38, 2022, doi: 10.19184/isj.v7i1.30901.
- [3] R. Takdirillah, "Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan," *Edumatic J. Pendidik. Inform.*, vol. 4, no. 1, pp. 37–46, 2020, doi: 10.29408/edumatic.v4i1.2081.
- [4] M. Badrul, "Algoritma Asosiasi Dengan Algoritma Apriori Untuk Analisa Data Penjualan," *J. Pilar Nusa Mandiri*, vol. XII, no. 2, pp. 121–129, 2016, [Online]. Available: <http://ejournal.nusamandiri.ac.id/index.php/pilar/article/view/266>
- [5] I. A. Ashari, A. Wirasto, and D. N. Triwibowo, "Implementasi Market Basket Analysis dengan Algoritma Apriori untuk Analisis Pendapatan Usaha Retail Implementation of Market Basket Analysis with Apriori Algorithm for Retail Business Income Analysis," vol. 21, no. 3, 2022, doi: 10.30812/matrik.v21i3.1439.
- [6] N. Anwar, F. Adikara, R. Setiyati, R. Satria, and A. Satriawan, "Data Mining Menggunakan Metode Algoritma Apriori Pada Vending Machine Product Display," *JBASE - J. Bus. Audit Inf. Syst.*, vol. 4, no. 2, pp. 23–31, 2021, doi: 10.30813/jbase.v4i2.3004.
- [7] A. Triayudi, "Penerapan Algoritma Apriori Data Mining Untuk Menentukan Penyusunan Layout Barang Pada Toko Ritel," vol. 4, no. 2, pp. 1123–1128, 2022, doi: 10.47065/bits.v4i2.2303.
- [8] J. G. Susanto and S. Budi, "Penerapan Data Science Pada Dataset Pokemon," vol. 4, no. November, pp. 243–254, 2022.
- [9] A. F. Lestari and M. Hafiz, "Penerapan Algoritma Apriori Pada Data Penjualan Barbar Warehouse," *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 1, p. 96, 2020, doi: 10.35314/isi.v5i1.1317.
- [10] Y. Andini et al., "Penerapan Data Mining Terhadap Tata Letak Buku," vol. XI, no. 1, pp. 9–15, 2022.
- [11] E. Umar, D. Manongga, and A. Iriani, "Market Basket Analysis Menggunakan Association Rule dan Algoritma Apriori Pada Produk Penjualan Barang," vol. 6, pp. 1367–1377, 2022, doi: 10.30865/mib.v6i3.4217.
- [12] W. F. Fitriana Harahap\*, Nidia Enjelita Saragih, Ester Dayanti Paskaria Situmeang, Elida Tuti, Erwin Ginting, "Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori," *J. Media Inform. Budidarma*, vol. 6, no. 2, pp. 1159–1165, 2022, doi: 10.30865/mib.v6i2.3937.



# LoA Jurnal Kesatria

The screenshot shows the homepage of the KESATRIA journal. At the top, there is a green banner with the journal's logo (a stylized gear and head profile) and the title "KESATRIA". Below the banner, there is a navigation menu with links for Home, About, Login, Register, Search, Current, and Archives. A breadcrumb navigation "Home > Kesatria : Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)" is displayed. The main content area features a large image of a book cover for "KESATRIA" and a bar chart showing citation statistics from Google Scholar. The chart has the following data:

Tahun	Citation
2019	34
2020	34
2021	73
2022	36
2023	36

Below the chart, there is a section titled "Citation Statistics by Author My ID" with links to AIM and Scope, Indexing & Abstracting, Author Guidelines, Publication Ethics, Access Submission, and Submission Guidelines.

**Journal Information:**

- Journal title: KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)
- Initials: KESATRIA
- Frequency: 4 issues per year (January, April, July and October)

The screenshot shows a submission page for the KESATRIA journal. The header includes the journal logo and title. The page contains the following text:

Pematangsiantar, 27 Maret 2023  
No : 017/SRT-LOA/KESATRIAVOL4NO2/III/2023  
Hal : Surat Penerimaan Naskah Publikasi Artikel Ilmiah

**Kepada Yth:**  
Bapak/Ibu Penulis (*Author*)  
**Mochammad Septa Sandy, Hamzah Setiawan, Uce Indahyanti**  
Di  
Tempat  
Assalamualaikum Wr. Wb  
Salam Sejahtera  
Terimakasih telah mengirimkan artikel ilmiah untuk diterbitkan pada KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen), ISSN Online : 2720-992X dengan Judul:

**Analisis Data Mining Produk Retail Menggunakan Metode Asosiasi Dengan Menerapkan Algoritma Apriori**

Berdasarkan hasil *review*, artikel tersebut dinyatakan **DITERIMA** untuk dipublikasikan di Jurnal kami **Volume 4, Nomor 2, April 2023**. Kami akan mengirimkan *softcopy* edisi tersebut pada akhir bulan penerbitan ke email penulis. Artikel tersedia secara online di <https://tunasbangsa.ac.id/pkm/index.php/kesatria>.

