



DATA MINING

IMPLEMENTASI ALGORITMA APRIORI
DAN FP_GROWTH

Lismardiana Sembiring, ST., M.Kom
Dr. Thamrin, BSc (Hons), MM
Wanra Tarigan, ST., M.Kom



Biografi



Lismardiana Sembiring, ST., M.Kom, Lahir di Lau Timah pada Tanggal 03 Oktober 1978, S1 Teknik Informatika dari Universitas Harapan Medan pada Tahun 2005. S2 Teknik Informatika dari Universitas Sumatera Utara Tahun 2015. Aktif Megajar dan Sebagai Dosen Tetap di Universitas Mandiri Bina Prestasi Medan. Aktif Penulis Jurnal Nasional dan Internasional.



Dr. Thamrin, BSc (Hons), MM was born on 1st January 1977, is a lecturer at Institut Bisnis Informasi Teknologi dan Bisnis in Medan. He graduated from Staffordshire University in Bachelor of IT and Doctorate in Management from Universitas Pasundan. Currently he is the Head of Information Systems study program at Institut Bisnis Informasi Teknologi dan Bisnis. He has been teaching in IT field for 18 years and also active as an experienced software developer for more than 20 years.
(Google Scholar ID: 8gzgV5gAAAAJ ; Sinta ID : 6764796)



Wanra Tarigan, ST., M.Kom, Lahir di Lau Meciho (Dairi), 12 Februari 1971, Lulus S1 Program Studi Teknik Informatika Fakultas Teknik Universitas Satya Negara Indonesia Jakarta Tahun 1997, Lulus S2 Program Studi Ilmu Komputer Universitas Gadjah Mada Yogyakarta Tahun 2003. Saat ini aktif sebagai Dosen di Program Studi Sistem Informasi Universitas Mandiri Bina Prestasi. Sebagai Dosen sejak 1997 dalam mata kuliah Manajemen Proyek Sistem Informasi, Rekayasa Perangkat Lunak dan Pemrograman Basis Data, sejak 2015 telah mendapatkan sertifikat sebagai pendidik (sertifikasi Dosen).



☎ 0858 5343 1992
✉ eurekaediaaksara@gmail.com
📍 Jl. Banjaran RT.20 RW.10
Bojongsari - Purbalingga 53362



DATA MINING
IMPLEMENTASI ALGORITMA APRIORI DAN FP_GROWTH

Lismardiana Sembiring, ST.,M.Kom
Dr. Thamrin, BSc (Hons), MM
Wanra Tarigan, ST.,M.Kom



eureka
media aksara

PENERBIT CV.EUREKA MEDIA AKSARA

**DATA MINING
IMPLEMENTASI ALGORITMA APRIORI DAN FP_GROWTH**

Penulis : Lismardiana Sembiring, ST.,M.Kom
Dr. Thamrin, BSc (Hons), MM
Wanra Tarigan, ST.,M.Kom

Desain Sampul : Eri Setiawan

Tata Letak : Siwi Rimayani Oktora

ISBN : 978-623-5382-47-0

Diterbitkan oleh : **EUREKA MEDIA AKSARA, MEI 2022**
ANGGOTA IKAPI JAWA TENGAH
NO. 225/JTE/2021

Redaksi:

Jalan Banjaran, Desa Banjaran RT 20 RW 10 Kecamatan Bojongsari
Kabupaten Purbalingga Telp. 0858-5343-1992

Surel : eurekamediaaksara@gmail.com

Cetakan Pertama : 2022

All right reserved

Hak Cipta dilindungi undang-undang

Dilarang memperbanyak atau memindahkan sebagian atau seluruh isi buku ini dalam bentuk apapun dan dengan cara apapun, termasuk memfotokopi, merekam, atau dengan teknik perekaman lainnya tanpa seizin tertulis dari penerbit.

KATA PENGANTAR

Puji Syukur saya ucapkan kepada TYME atas Berkah, Rahmat, Kesehatan serta perlindungan yang diberikanNYA kepada penulis, sehingga penulis dapat merampungkan buku yang berjudul : **“DATA MINING IMPLEMENTASI ALGORITMA APRIORI DAN FP_GROWTH”**. terimakasih kasih juga buat keluarga kecilku suami dan anakku yang selalu mensupport penulis setiap harinya. Dalam pengerjaan buku ini yang dikerjakan selama dalam kurun waktu 1 tahun lamanya didasari dengan pengalaman penulis selama melakukan penelitian terutama pada waktu penulis menyelesaikan studi S2 ilmu Komputer serta pengalaman penulis dalam mempublikasikan karya ilmiah dan jurnal jurnal baik jurnal nasional bereputasi seperti sinta dan scopus maupun jurnal internasional bereputasi. Di dalam buku ini kemungkinan ada kehilafan atau kesalahan dalam melakukan penulisan baik secara teoritis maupun secara lisan saya selaku penulis buku tersebut mohon maaf dan mohon saran dan kritiknya yang sifatnya membangun dan positif sangat saya harapkan. Dengan adanya buku ini sekiranya dapat menjadi peluang atau ide ide yang lebih menarik lagi bagi pembaca buku ini. Adapun buku yang sudah di terbitkan ini bertujuan untuk menjadi inspirasi bagi pembaca dalam melakukan Penelitian, Pengajaran, sebagai alat bantu dalam memecahkan masalah dalam pengambilan keputusan dan dapat juga dijadikan sebagai metode /konsep/Pedoman Prbolem Solving di sebuah Perusahaan Industri kecil/besar.

DAFTAR ISI

KATA PENGANTAR.....	iii
DAFTAR ISI	iv
DAFTAR TABEL	vi
DAFTAR GAMBAR	ix
BAB 1 PENDAHULUAN.....	1
1.1. Data Mining	1
1.2. Fungsi Data Mining.....	2
1.3. Metode Data Mining	3
1.3.1. Tahapan Data Mining	3
1.4. Tujuan Data Mining	8
1.5. Penerapan Data Mining.....	9
BAB 2 ALGORITMA APRIORI	12
2.1. Algoritma Apriori	12
2.2. Studi Kasus Bagian I	15
2.3. Association Rule Mining.....	22
BAB 3 PENERAPAN ALGORITMA APRIORI.....	24
3.1. Data Mining untuk Data yang Besar	24
3.2. Cleaning Database TransactionID_salesdataset 250 records	26
3.3. Cleaning Database TransactionID_salesdataset 500 records	30
3.4. TransactionID_sales dataset 250-500 records Menggunakan Apriori	53
3.5. Pengujian Dataset 1000 records Menggunakan Algoritma Apriori	57
BAB 4 PENERAPAN ALGORITMA FP_GROWTH PADA DATA MINING	80
4.1. FP_Growth.....	80
4.2. Contoh Studi Kasus bagian ke II.....	81
4.3. Frequent Itemset dataset 500 records menggunakan FP_Tree.....	84
4.4. Hasil Pengujian dataset1000 menggunakan FP- Growth	87
BAB 5 KESIMPULAN	94

DAFTAR PUSTAKA.....	95
TENTANG PENULIS	97

DAFTAR TABEL

Tabel 2.1 Dataset yang digunakan	16
Tabel 2.2 Data Frequent itemset 37 Records.....	17
Tabel 2.3 K-1 ItemSet.....	17
Tabel 2.4 K-2 Frequent ItemSet.....	18
Tabel 2.5 K-3 Frequent ItemSet.....	19
Tabel 2.6 K-4 ItemSet.....	20
Tabel 2.7 Asosiasi Rule Mining	20
Tabel 2.8 Keterangan Asosiasi Rule Mining.....	21
Tabel 3.1 Data Spesifikasi	26
Tabel 3.2 Celaning Database city transactionID_sales Dataset 250 Records	26
Tabel 3.3 Cleaning Database state transactionID_sales Dataset 250 Records.....	27
Tabel 3.4 Cleaning Database country transactionID_sales Dataset 250 records	29
Tabel 3.5 Hasil Pengkodean dari Filter Data transaksiID_sales 250 records untuk city, state dan country	30
Tabel 3.6 Cleaning Database city transactionID_sales dataset 500 records.....	30
Tabel 3.7 Data processing state transactionID_sales dataset 500 records.....	32
Tabel 3.8 Cleaning Database Country transactionID_sales dataset 500 records.....	35
Tabel 3.9 Hasil Pengkodean dari Filter Data transaksiID_sales 500 records untuk city,state dan country	36
Tabel 3.10 Cleaning Database city transactionID_Sales Dataset 750 records.....	37
Tabel 3.11 Cleaning Database state transactionID_Sales Dataset 750 records.....	39
Tabel 3.12 Cleaning Database country transactionID_Sales Dataset 750 records	42
Tabel 3.13 Hasil Pengkodean dari Filter Data transaksiID_sales 750 records untuk city,state dan country	43

Tabel 3.14 Cleaning Database city TransactionID_Sales Dataset 1000 records.....	44
Tabel 3.15 Cleaning Database state transactionID_Sales Dataset 1000 records.....	47
Tabel 3.16 Cleaning Database Country transactionID_Sales Dataset 1000 records.....	51
Tabel 3.17 Hasil Pengkodean dari Filter Data transaksiID_sales 1000 records untuk city,state dan country.....	53
Tabel 3.18 TransactionID_sales Dataset250-500 records.....	53
Tabel 3.19 K-1 Frequent Item	54
Tabel 3.20 K-2 Frequent Itemset.....	54
Tabel 3.21 K-3 Frequent Itemset.....	55
Tabel 3. 22 K-4 Frequent Itemset.....	56
Tabel 3.23 Asosiasi Rule Mining.....	56
Tabel 3.24 TransactionID_sales Dataset1000 records	57
Tabel 3.25 K-1 Frequent Item	57
Tabel 3.26 K-2 Frequent Itemset.....	58
Tabel 3.27 K-3 Frequent Itemset.....	62
Tabel 3.28 K-4 Frequent Itemset.....	65
Tabel 3.29 K-5 Frequent Itemset.....	67
Tabel 3.30 K-6 Frequent Itemset.....	69
Tabel 3.31 Asosiasi Rule Mining.....	71
Tabel 3.32 Keterangan Simbol A.....	71
Tabel 3.33 Keterangan Simbol B	74
Tabel 3.34 Keterangan Simbol C.....	78
Tabel 4.1 TID Records.....	81
Tabel 4.2 Generate Frequent Itemset 2	82
Tabel 4.3 Assosiasi Rule Mining	83
Tabel 4.4 List Asosiasi Rule Mining	84
Tabel 4.5 TID_Ordered FP_Tree	85
Tabel 4.6 k-1 item	85
Tabel 4.7 Conditional Patternbase dataset 500	87
Tabel 4.8 TransactionID_sales Dataset 1000	87
Tabel 4.9 TID.Ordered Dataset 1000.....	88
Tabel 4.10 K-1 Frequent Item	88

Tabel 4.11 Tabel Conditional Pattern Base FP-Tree dataset 1000records	91
Tabel 4.12 Perbandingan waktu Penyelesaian Apriori + FP Growth	92
Tabel 4. 13 FP_Growth VS Apriori.....	93

DAFTAR GAMBAR

Gambar 1.1 Tahapan Data Mining (Han 2006).....	4
Gambar 2.1 Proses Iterasi Asosiasi Rule Mining Pada Algoritma Apriori.....	14
Gambar 2.2 Deskripsi Algoritma Apriori	15
Gambar 3.1 Data .csv file	25
Gambar 4.1 FP_Tree.....	83
Gambar 4.2 Fp-tree dataset500.....	86
Gambar 4.3 FP-Tree dataset 1000.....	90
Gambar 4.4 Grafik Perbandingan waktu Penyelesaian Apriori+ FP Growth.....	92



DATA MINING
IMPLEMENTASI ALGORITMA APRIORI VS FP_GROWTH



BAB

1

PENDAHULUAN

1.1. Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. (Gupta 2014)

Pengertian Data Mining menurut para ahli: (Michael J.A Berry 2004)

Berry “ *data mining adalah sebuah aktivitas analisa yang memiliki jumlah besar demi menemukan pola (pattern) dan aturan (rule) yang berguna*”

Data mining adalah proses kegiatan dalam menemukan pola yang menarik dari suatu data dalam jumlah besar , data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya. Data mining dapat juga dihubungkan dengan bidang ilmu lainnya, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysis, image database, signal processing (Han, 2006).

Berdasarkan beberapa pengertian tersebut dapat diartikan bahwa data mining adalah teknik menggali data untuk dijadikan informasi bernilai yang terpendam atau

BAB

2

ALGORITMA APRIORI

2.1. Algoritma Apriori

Algoritma Apriori termasuk jenis aturan asosiasi pada data mining atau market basket analysis. Algoritma apriori salah satu algoritma yang berfungsi sebagai untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah item yang sering muncul pada sebuah datamining dan memiliki frekuensi support diatas ambang tertentu yang disebut dengan istilah minimum support dan minimum Confidence. Pola frekuensi tinggi ini digunakan untuk menyusun aturan asosiasi dan juga beberapa teknik data mining. Dari teknik proses pengerjaannya algoritma apriori dibagi menjadi beberapa tahap yang disebut iterasi. Tiap iterasi menghasilkan pola frekuensi tinggi dengan panjang yang sama dimulai dari iterasi pertama yang menghasilkan pola frekuensi tinggi dengan panjang satu. Di iterasi pertama ini support dari setiap item dihitung dengan membaca database. Setelah support dipilih sebagai pola frekuensi tinggi dengan panjang 1 atau sering disebut 1-item set. Singkatan k-itemset berarti satu set yang dari k item. (Yuan, Itwee, and Wei 2017)

Untuk asosiasi rule mining merupakan prosedur untuk mencari hubungan antar item dalam suatu data set yang ditentukan meliputi: mencari kombinasi yang sering terjadi dari suatu item set dan mendefenisikan condition dan result (*untuk conditional association rule*). Dalam menentukan

BAB 3

PENERAPAN ALGORITMA APRIORI PADA DATA MINING

3.1. Data Mining untuk Data yang Besar

Pada bab ini, terfokus pada penerapan algoritma Apriori dan proses iterasi iterasi yang dilakukan sehingga terjadinya Frequent Item/itemset. Dari konsep algoritma apriori melakukan proses *frequent itemset* untuk menemukan aturan asosiasi antara suatu kombinasi *item/itemset*. *Frequent item/itemset* akan terus dilakukan mulai dari generasi 1{k-1, k-2, k-3,...k-n} *item/itemset* sampai tidak dapat di generasikan lagi.

Dalam melakukan pengujian data dalam kapasitas yang besar bersumber dari: <https://support.spatialkey.com/spatialkey-sample-csv-data/datasales> data yang berisikan database transaksi penjualan sales di suatu wilayah, database telah pernah dilakukan penelitian apriori yang berformat .CSV. Data tersebut akan digunakan sebagai data pelatihan dan juga data pengujian. Data yang digunakan dalam proses ini adalah sebagai berikut :

1. Data yang digunakan adalah file database berisikan Transaction sales penjualan product di seluruh negara dengan format. CSV
2. Banyak data yang digunakan adalah ± 1500 records.

Berdasarkan data yang ada, data yang akan digunakan untuk penelitian ini mempunyai atribut: City, State, Country

BAB 4

PENERAPAN ALGORITMA FP_ GROWTH/FP_TREE PADA DATA MINING

4.1. FP_Growth

Algoritma FP-Growth merupakan pengembangan dari algoritma Apriori. Algoritma Frequent Pattern Growth adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (frequent itemset) dalam sebuah kumpulan data. Langkah langkah penyelesaian FP_Growth

1. Penyiapan Dataset.
2. Pencarian Frequent Itemset (Item yang sering muncul)
3. Dataset diurutkan Berdasarkan Priority.
4. Pembuatan **FP-Tree** Berdasarkan Item yang sudah diurutkan.
5. Pembangkitan Conditional Pattern Base.
6. Pembangkitan Conditional **FP-tree**.
7. Pembangkitan **Frequent Pattern**.
8. Mencari Support.

BAB

5

KESIMPULAN

Berdasarkan pengujian data yang dilakukan pada bab 3 dan bab 4 terhadap algoritma apriori dan algoritma FP_Growth / FP_Tree dengan studi kasus yang sama maka dapat disimpulkan sbb:

1. Frequent Item/Itemset jauh lebih sedikit yang dihasilkan
2. Lebih efisien dalam penggunaan waktu
3. Tahapan penyelesaian lebih cepat, disebabkan tanpa melakukan candidate generation
4. Frequent Pattern item/itemset yang bernilai 1, tidak dibuang sedangkan di FP_Tree dihilangkan sehingga model frequent pattern item/itemset lebih jelas lagi.
5. Penganalisaan terhadap frequent item/itemset lebih mudah dipahami disebabkan frequent item/itemset lebih mudah dipahami disebabkan frequent item/itemset lebih sedikit yang dihasilkan.
6. Selain melakukan perbandingan dengan algoritma apriori dan FP_Tree pada pengujian data mining ini dapat mengukur standart deviasi error pada apriori dan FP_Tree perlu di kembangkan lagi.

Mungkin setelah membaca buku ini dapat menjadi inspirasi bagi pembaca, penulis buku dan peneliti dalam melakukan penelitian dan mengembangkan algoritma apriori dan FP_Tree lebih tajam dan luas lagi sehingga dapat dijadikan sebagai kanal diskusi dan penelitian untuk lebih lanjut lagi.

DAFTAR PUSTAKA

- Abbass, Hussein A., Ruhul A. Sarker, and Charles S. (Charles Sinclair) Newton. 2002. *Data Mining: A Heuristic Approach*. Idea Group.
- Agarwal, Shivam. 2014. *Data Mining: Data Mining Concepts and Techniques. Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*. <https://doi.org/10.1109/ICMIRA.2013.45>.
- Gupta, Richa. 2014. "Journey from Data Mining to Web Mining to Big Data." *International Journal of Computer Trends and Technology* 10 (1). <http://www.ijcttjournal.orgpage18>.
- Kantardzic, Mehmed. 2011. *Data Mining Data Mining Concept, Models, Methods, and Algorithms*.
- Leskovec, Jure, Anand Rajaraman, and Jeffrey D Ullman. 2010. "Mining of Massive Datasets."
- Li, Yuanyuan, and Shaohong Yin. 2019. "Mining Algorithm for Weighted FP-Growth Frequent Item Sets Based on Ordered FP-Tree." *International Journal of Engineering and Management Research* 09 (05): 154-58. <https://doi.org/10.31033/ijemr.9.5.22>.
- Lismardiana. 2019. "ASSOSIASI RULE MINING POLA PEMETAAN MUTU PENDIDIKAN SEKOLAH SD NEGERI DAN SD SWASTA DI KECAMATAN MEDAN KOTA MENGGUNAKAN ALGORITMA APRIORI." *Jurnal Pembangunan Perkotaan* 7 (1). <http://ejpp.balitbang.pemkomedan.go.id/index.php/JPP>.
- Michael J.A Berry, Gordon S. Linoff. 2004. *Data Mining Techniques: For Marketing, Sales, and Customer Relationship*. Second Edi. Wiley Publishing, Inc.

- Ng, Wing Shui. 2016. "Web Data Mining in Education: Decision Support by Learning Analytics with Bloom's Taxonomy." *Web Data Mining and the Development of Knowledge-Based Decision Support Systems* 7 (6): 58-77. <https://doi.org/10.4018/978-1-5225-1877-8.ch005>.
- Yuan, Ling, Wasan Itwee, and Shikang Wei. 2017. "Association Rule Mining Technique with Optimized FP-Tree Algorithm." *INTERNATIONAL JOURNAL OF ADVNACED STUDIES IN COMPUTER SCIENCE AND ENGINEERING IJASCSE VOLUME*. Vol. 6.

TENTANG PENULIS

Lismardiana, ST.,M.Kom



Lismardiana, ST.,M.Kom, Lahir di Lau Timah pada Tanggal 03 Oktober 1978 , S1 Teknik Informatika dari Universitas Harapan Medan pada Tahun 2005. S2 Teknik Informatika dari Universitas Sumatera Utara Tahun 2015. Aktif Megajar dan Sebagai Dosen Tetap di Universitas Mandiri Bina Prestasi Medan. Aktif Penulis Jurnal Nasional dan Internasioal.

Dr. Thamrin, BSc (Hons), MM



Dr. Thamrin, BSc (Hons), MM was born on 1st January 1977, is a lecturer at *Institut Bisnis Informasi Teknologi dan Bisnis* in Medan. He graduated from Staffordshire University in Bachelor of IT and Doctorate in Management from *Universitas Pasundan*. Currently he is the Head of Information Systems study program at *Institut Bisnis Informasi Teknologi dan Bisnis*. He has been teaching in IT field for 18 years and also active as an experienced software developer for more than 20 years.

(Google Scholar ID: 8gzgV5gAAAAJ ; Sinta ID : 6764796)

Wanra Tarigan,ST.,M.Kom



Wanra Tarigan,ST.,M.Kom, Lahir di Lau Meciho (Dairi), 12 Februari 1971, Lulus S1 Program Studi Teknik Informatika Fakultas Teknik Universitas Satya Negara Indonesia Jakarta Tahun 1997, Lulus S2 Program Studi Ilmu Komputer Universitas Gadjah Mada Yogyakarta Tahun 2003. Saat ini aktif sebagai Dosen di Program Studi Sistem Informasi Universitas Mandiri Bina Prestasi. Sebagai Dosen sejak 1997 dalam mata kuliah Manajemen Proyek Sistem Informasi, Rekayasa Perangkat Lunak dan Pemrograman Basis Data, sejak 2015 telah mendapatkan sertifikat sebagai pendidik (sertifikasi Dosen)