

KAJIAN LITERATUR MASALAH – MASALAH YANG DIHADAPI DENGAN MENGGUNAKAN ALGORITMA FP-GROWTH DAN APRIORI

LITERATURE STUDY PROBLEMS FACED USING FP-GROWTH AND APRIORI ALGORITHMS

Sabrina Aulia Rahmah^{1*}, Zulham², Ibnu Rusydi³

Universitas Dharmawangsa^{1,2,3}

sabrinaaulia@dharmawangsa.ac.id^{1*}, zulham@dharmawangsa.ac.id²,

ibnurusydi@dharmawangsa.ac.id

Abstrak

Data Mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata dengan melakukan penggalan pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basisdata. Artikel ini akan membahas masalah-masalah bagaimana dalam menentukan suatu pola yang saling berhubungan sehingga dengan keterkaitan tersebut. Berdasarkan tujuan dari algoritma FP-Growth dan apriori dapat memaksimalkan solusi suatu masalah yang sering muncul di masa yang akan datang.

Kata kunci: Data Mining, FP-Growth dan Apriori

Abstract

Data Mining is a series of processes to explore added value in the form of information that has not been known manually from a database by extracting patterns from data with the aim of manipulating data into more valuable information obtained by extracting and recognizing important patterns or pulls from the data contained in the database. This article will discuss the problems of how to determine a pattern that is interconnected so that the relationship is. Based on the objectives of the FP-Growth algorithm and a priori, it can maximize the solution to a problem that often arises in the future.

Keywords: Data Mining, FP-Growth dan Apriori

1. Pendahuluan

Perkembangan teknologi saat ini yang semakin maju dalam segala bidang menuntut manusia untuk menemukan suatu strategi yang dapat meningkatkan kinerja dan kualitas dalam hal apapun. Di dalam teknologi tersebut yang menyentuh berbagai aspek kehidupan dituntut untuk bisa bekerja lebih efisien agar hasil yang diraih bisa langsung dirasakan oleh orang yang membutuhkannya. Pada umumnya pendekatan dengan algoritma menghasilkan sesuatu yang nantinya bisa mengubah basis data sehingga pada akhirnya tercipta sesuatu yang menentukan sebuah pola yang diinginkan peneliti. Dalam hal ini peneliti melakukan pendekatan dengan menggunakan algoritma FP-Growth dan Apriori yang mana mencari hubungan atau keterkaitan dalam suatu masalah. Untuk memaksimalkan hasil penelitian diperlukan pengelompokan basis data agar dalam mencari hubungan atau *association rules* agar yang dihasilkan lebih efektif dan efisien.

Association rules adalah suatu metode data mining yang bertujuan untuk mencari sekumpulan items yang sering muncul bersamaan. Umumnya association rule ini dianalogikan dengan keranjang belanjaan. Dari keranjang belanjaan para pengunjung supermarket akan dapat diketahui, barang apa saja yang sering dibeli bersamaan dan barang mana saja yang tidak. Pendekatan dengan menggunakan algoritma FP-Growth dan Apriori sangat dibutuhkan dalam dunia usaha maupun aspek lainnya. Karakteristik algoritma FP-Growth merupakan struktur data yang menggunakan tree atau disebut

dengan FP-Tree. Dengan menggunakan FP-Tree, algoritma FP- Growth dapat langsung mengekstrak frequent Itemset dari FP-Tree. Penggalan itemset yang frequent dengan menggunakan algoritma FP-Growth akan dilakukan dengan cara membangkitkan struktur data tree atau disebut dengan FP- Tree. Metode FP-Growth dapat dibagi menjadi 3 tahapan utama yaitu sebagai:

1. Tahap pembangkitan conditional pattern base,
2. Tahap pembangkitan conditional FP-Tree, dan
3. Tahap pencarian frequent itemset.

Dengan begitu setiap permasalahan yang muncul bersamaan ataupun yang mempunyai hubungan akan dapat dideteksi. Adapun FP- Tree adalah sebuah pohon dengan definisi sebagai berikut:

1. FP-Tree dibentuk oleh sebuah akar yang diberi label null, sekumpulan sub-tree yang beranggotakan item-item tertentu, dan sebuah tabel frequent header.
2. Setiap simpul dalam FP-Tree mengandung tiga informasi penting, yaitu label item, menginformasikan jenis item yang direpresentasikan simpul tersebut, support count, merepresentasikan jumlah lintasan transaksi yang melalui simpul tersebut, dan pointer penghubung yang menghubungkan simpul-simpul dengan label item sama antar-lintasan, ditandai dengan garis panah putus-putus.

Dari hal di atas setiap masalah yang mempunyai keterkaitan data akan dapat terdistribusi kedalam suatu basis data sehingga masalah tersebut dapat diatasi lebih cepat dan efisien. Untuk itu data mining juga banyak digunakan dalam mengklasifikasikan kriteria maupun itemset yang sering muncul secara bersamaan. Pengembangan - pengembangan yang dilakukan pada algoritma FP-Growth dan apriori perlu dilakukan untuk menyesuaikan segala kebutuhan yang nantinya diperlukan. Peningkatan dalam penggunaan kedua algoritma tersebut untuk menghasilkan yang lebih baik di masa yang akan datang. Peningkatan-peningkatan yang terkadang muncul pada saat melakukan penelitian dan terus berkembang menjadi suatu kemajuan teknologi diperlukan guna sebagai penunjang dunia usaha dan industri.

Oleh karena itu peneliti diharapkan bisa mendapatkan solusi yang efektif dari masalah-masalah yang sedang dihadapi pada saat sekarang ini. Adapun masalah-masalah tersebut diklasifikasikan sehingga secara teoritis menghasilkan solusi yang dinamis yang lebih kreatif dan inovatif.

2. Tinjauan pustaka dan pengembangan hipotesis

Tujuan melakukan pendekatan algoritma adalah untuk memudahkan dalam menentukan suatu pola yang dibutuhkan. Adapun studi pustaka yang dilakukan melalui review beberapa jurnal menghasilkan segala kelebihan dan kekurangan dari suatu permasalahan yang diselesaikan dengan menggunakan algoritma FP-Growth dan apriori pada data mining. Tingkat efektivitas dalam penggunaannya lebih mengarah kepada suatu pola yang akan digunakan pada masa mendatang.

3. Metode penelitian

Penelitian ini menggunakan metode literatur dengan melakukan review 100 jurnal yang berhubungan dengan algoritma FP-Growth dan apriori. Jurnal tersebut adalah 10 tahun terakhir. Adapun tinjauan sistematis yang dilakukan adalah berdasarkan data literatur dari penelitian sebelumnya, dilakukan berdasarkan beberapa langkah sebagai berikut :

1. Jumlah total studi awal yang direview ditemukan dari e-journal sebanyak 100.
2. Dilakukan pensortiran jurnal yang berhubungan dengan masalah yang dihadapi pada algoritma FP-Growth dan apriori sebanyak 41 jurnal.

4. Hasil dan pembahasan

Untuk mendapatkan hasil dan tujuan yang maksimal perlu dilakukan tahap-tahap sebagai berikut :

Tahap pertama : Pricise statement of the problem (pernyataan tepat terhadap permasalahan) Sebelum mengakses perangkat lunak data mining, seorang analis harus memiliki kejelasan perihal 'pertanyaan apa yang akan ingin dijawabnya'. Jika tidak ada formula yang tepat untuk problematika yang ada maka anda hanya akan membuang-buang dan uang dalam membuat solusinya.

Tahap kedua : Initial exploration Tahap ini dimulai dengan mempersiapkan data yang juga juga termasuk kedalam data mining “cleaning” (misalnya : mengidentifikasi dan menyikatkan data yang dikodekan salah), transformasi data, memilih subset record, data set, langkah awal seleksi. Mendeskripsikan dan memvisualisasikan data adalah kunci dari tahap ini.

Tahap tiga : Model building and validation Tahap ini melibatkan pertimbangan terhadap ragam permodelan dan memilih yang terbaik bagi performasi prediktif.

Tahap ke-empat : Deployment Memilih aplikasi yang tepat berikut permodelan untuk membuat (generate) prediksi. Selanjutnya kita kan melihat rincian perihal tahapan-tahapan data mining.

Selanjutnya melakukan pengelompokkan Data Mining menjadi beberapa kelompok sebagai berikut :

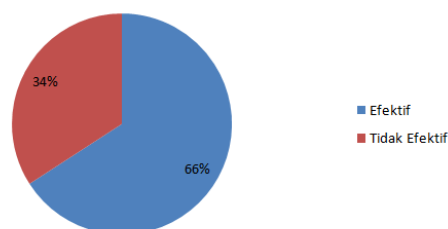
1. Deskripsi
Mencoba mencari cara bagaimana menggambarkan pola yang terdapat dalam data tersebut.
2. Estimasi
Mencari target yang nantinya akan dihasilkan sehingga mempunyai klasifikasi yang dibutuhkan dalam tahap berikutnya.
3. Prediksi
Menentukan hal-hal yang akan dihasilkan di masa mendatang agar penelitian sesuai dengan apa yang diharapkan.
4. Klasifikasi
Menentukan target variabel berdasarkan kategorinya.
5. Pengklusteran
Mengelompokkan hasil, pengamatan dan membentuk bagian-bagian dari objek yang mempunyai kemiripan.
6. Asosiasi
Mencari hubungan dari beberapa atribut yang selalu muncul bersamaan.

Dari klasifikasi jurnal yang direview dapat ditampilkan sebagai berikut :

Tabel 1 Judul Klasifikasi Jurnal

No	Algoritma	Jumlah
1	FP-Growth	41
2	Apriori-FP Growth	30
3	Kombinasi beberapa Algoritma	29

Dari tabel diatas didapatkan bahwa pembahasan mengenai penggunaan algoritma FP-Growth lebih banyak dibandingkan dengan algoritma lainnya. Dari tingkat efektivitas 41 jurnal yang sesuai didapat beberapa kategori sebagai berikut :



Gambar 1. Persentase penyelesaian masalah dengan algoritma FP-Growth

Dari persentase di atas didapati bahwasannya penyelesaian masalah dengan algoritma FP-Growth lebih efektif dibandingkan menggunakan algoritma apriori.

5. Kesimpulan

Dari hasil penelitian dan pembahasan dengan menggunakan algoritma FP-Growth lebih efisien dibandingkan algoritma apriori pada data mining karena FP-Growth memberikan pola yang lebih dinamis dan relevan dipakai pada masa yang mendatang.

Disamping itu juga FP-Growth dengan FP-Tree-nya memberikan hasil yang lebih kepada hubungan atau keterkaitan itemset yang muncul secara bersamaan. Dengan begitu peneliti bisa membuat solusi dalam penyelesaian masalah yang dihadapi.

REFERENSI

- [1] Bin Pei, Xiuzhen Wang and Fenmei Wang, "Parallelization of FP-growth Algorithm for Mining Probabilistic Numerical Data based on MapReduce", 2016 9th International Symposium on Computational Intelligence and Design.
- [2] K.Dharmarajan and M.A. Dorairangaswamy "Analysis of FP-Growth and Apriori Algorithms on Pattern Discovery from Weblog Data", 2016 IEEE International Conference on Advances in Computer Applications (ICACA).
- [3] DENG Lingling and LOU Yuansheng "Improvement and research of FP-growth algorithm based on distributed spark" 2015 International Conference on Cloud Computing and Big Data.
- [4] Hongwei Chen, Qixing Luo, Zexi Chen and Yingzhe Chen, "Distributed pruning optimization oriented FP-Growth method based on PSO algorithm", IEEE, 2017.
- [5] Avadh Kishor Singh, Ajeet Kumar and Ashish K. Maurya, "An Empirical Analysis and Comparison of Apriori and FP- Growth Algorithm for Frequent Pattern Mining", 2014 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT).
- [6] Dea Delvia Arifin, Shaufiah and Moch. Arif Bijaksana, "Enhancing Spam Detection on Mobile Phone Short Message Service (SMS) Performance using FP-Growth and Naive Bayes Classifier", The 2016 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob).
- [7] Rini Angrainingsih, Nach Rowi Khoirudin and Haryono Setiadi, "Discovering Drugs Combination Pattern Using FPGrowth Algorithm', Proc. EECSI 2017, Yogyakarta, Indonesia, 19-21 September 2017.
- [8] Meng Zhao, Wanjun Yu, Wenjing Lu, Quan Liu and Jinxiao Li, "Chinese Document Keyword Extraction Algorithm based on FP-Growth", 2016 International Conference on Smart City and Systems Engineering.
- [9] Hong-Yi Chang, Jia-Chi Lin, Mei-Li Cheng and Shih-Chang Huang, "A Novel Incremental Data Mining Algorithm based on FP-Growth for Big Data", 2016 International Conference on Networking and Network Applications.
- [10] Omair Shafiq, Reda Alhadj and Jon G. Rokne, "Frequent Pattern Mining using Semantic FP-Growth for Effective Web Service Ranking", 2014 IEEE International Conference on Web Services.
- [11] Mungyu Lee and Junho Choi, "APT Attack Behavior Pattern Mining using the FP-Growth Algorithm", 2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC).
- [12] Vandana Rajput and Asst. Prof. Amit Manjhar, "Model for Forecasting Dengue Disease Using Genetic based Weighted FP-Growth", International Conference on Trends in Electronics and Informatics ICEI, 2017.
- [13] Chenlu Li and Xiangjun Dong, "FP-Growth Based Method for Mining Infrequent and Frequent Itemsets with 2-Level Minimum Support", 2016 5th International Conference on Computer Science and Network Technology (ICCSNT)
- [14] Xiaoting Wei, Yunlong Ma, Feng Zhang, Min Liu and Weiming Shen, "Incremental FP-Growth Mining Strategy for Dynamic Threshold Value and Database Based on MapReduce", Proceedings of the 2014 IEEE 18th International Conference on Computer Supported Cooperative Work in Design.
- [15] Pan Zhaopeng, Liu Peiyu and Yi Jing, "An Improved FP-tree Algorithm for Mining Maximal Frequent Patterns", 2018 10th International Conference on Measuring Technology and Mechatronics Automation.
- [16] Raja Varma Pamba and Raja Varma Pamba, "A Novice approach for web document clustering using FP Growth based Fuzzy Particle Swarm Optimization", 2016 3rd International Conference on Soft Computing & Machine Intelligence.

- [17] Shivangee Agrawal and Vikas Sejwar, "Crime Identification using FP-Growth and Multi Objective Particle Swarm Optimization", International Conference on Trends in Electronics and Informatics, ICEI 2017.
- [18] Amna Tehreem, Sajid Gul Khawaja, Muhammad Usman Akram, Shoab A. Khan and Muhammad Ali, "Parallel Architecture for Implementation of Frequent Itemset Mining Using FP-Growth", 2017 International Conference on Signals and Systems (ICSigSys).
- [19] Vaishali Patil, Ramesh Vasappanavara and Tushar Ghorpade, "Securing in association rule mining using secure sum Technique with FP growth algorithm in horizontally partitioned database", 2016 International Conference on Control, Computing, Communication and Materials (ICCCCM).
- [20] Shengbo Jiang, Xueli Pan, Yi Xiao, Qiangming Zhou, Miao Li, and Jun Yang, "Identification of vulnerable lines in large power grid based on FP-growth algorithm", The 6th International Conference on Renewable Power Generation (RPG), J. Eng., 2017, Vol. 2017, Iss. 13, pp. 1862–1866.
- [21] Jiye Wang and Zhihua Cheng, "FP-Growth based Regular Behaviors Auditing in Electric Management Information System", Procedia Computer Science 139 (2018) 275–279.
- [22] Meera Narvekar and Shafaque Fatma Syed, "An optimized algorithm for association rule mining using FP tree", International Conference on Advanced Computing Technologies and Applications (ICACTA-2015), Procedia Computer Science 45 (2015) 101 – 110.
- [23] Shivam Sidhu, Upendra Kumar Meena, Aditya Nawani, Himanshu Gupta and Narina Thakur, "FP Growth Algorithm Implementation", International Journal of Computer Applications (0975 – 8887) Volume 93 – No.8, May 2014.
- [24] M.S. Mythili and A.R. Mohamed Shanavas, Ph.D, "Performance Evaluation of Apriori and FP-Growth Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 79 – No.10, October 2013.
- [25] M.Kavitha and S.T.Tamil Selvi, "Comparative Study on Apriori Algorithm and Fp Growth Algorithm with Pros and Cons", International Journal of Computer Science Trends and Technology (IJCS T) – Volume 4 Issue 4, Jul - Aug 2016.
- [26] Rakesh Kumar Soni, Prof. Neetesh Gupta and Prof. Amit Sinhal, "An FP-Growth Approach to Mining Association Rules", International Journal of Computer Science and Mobile Computing, IJCSMC, Vol. 2, Issue. 2, February 2013, pg.1 – 5.
- [27] Abdullah Saad Almalaise Alghamdi, "Efficient Implementation of FP Growth Algorithm-Data Mining on Medical Data", IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.12, December 2011.
- [28] Abdulkader M. Al-Badani and Basheer M. Al-Maqaleh, "Efficient Mining of Frequent Itemsets using Improved FP-Growth Algorithm", International Journal of Applied Information Systems (IJ AIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA - Volume 12 – No. 14, July 2018 – www.ijais.org.
- [29] Rizky Fitria, Warnia Nengsih, dan Dini Hidayatul Qudsi, "Implementasi Algoritma FP-Growth dalam penentuan pola hubungan kecelakaan lalu lintas", Jurnal Sistem Informasi (Journal of Information Systems). 2/13 (2017), 118-124- DOI: <http://dx.doi.org/10.21609/jsi.v13i2.551>.
- [30] Alfannisa Annurullah Fajrin and Algifanri Maulana, "Penerapan Data Mining untuk Analisis Pola Pembelian Konsumen dengan Algoritma Fp-Growth pada Data Transaksi Penjualan Spare Part Motor", Kumpulan jurnaL Ilmu Komputer (KLIK), Volume 05, No.01 Februari 2018, ISSN: 2406-7857.
- [31] Zunera Farooq and Professor Vinod Sharma, "Association Rule Mining for Modelling Academic Resources using FP Growth Algorithm", International Journal of Computer Systems (ISSN: 2394-1065), Volume 03– Issue 05, May, 201, <http://www.ijconline.com/>.
- [32] Amanvir Kaur and Dr. Gagandeep Jagdev, "Analyzing Working of FP-Growth Algorithm for Frequent Pattern Mining", International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), Volume 4, Issue 4, 2017, PP 22-30, ISSN 2349-4840 (Print) & ISSN 2349-4859 (Online).
- [33] Yi Zeng, Shiqun Yin, Jiangyue Liu, and Miao Zhang, "Research of Improved FP-Growth Algorithm in Association Rules Mining", Hindawi Publishing Corporation, Scientific

- Programming, Volume 2015, Article ID 910281, 6 pages, <http://dx.doi.org/10.1155/2015/910281>.
- [34] Bhavesh V. Berani and Dr.Chirag Thaker,"FP Growth Algorithm for finding patterns in Semantic Web", 2014 IJEDR | Volume 2, Issue 4 | ISSN: 2321-9939.
- [35] Princy Shrivastava, Ankita Hundet, Babita Pathik and Shiv Kumar,"A Survey on FP (Growth) Tree Using Association rule Mining",International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Volume: 04 Issue: 07 - July -2017, www.irjet.net p-ISSN: 2395-0072.
- [36] Aroosa Hameed,"FP Growth Tree Implementation In Bank Transactional Databases",ITEE Journal (Information Technology & Electrical Engineering), Volume 5, Issue 3, June 2015.
- [37] Garima Sinha and Dr. S. M.Ghosh,"Identification of Best Algorithm in Association Rule Mining Based on Performance", IJCSMC, Vol. 3, Issue. 11, November 2014, pg.38– 45.
- [38] M.Sinthujal, Dr. N. Puviarasan and Dr. P.Aruna,"Research of Improved FP-Growth (IFP) Algorithm in Association Rules Mining", International Journal of Engineering Science Invention (IJESI) ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726, www.ijesi.org - PP. 24-31
- [39] Gagandeep Kaur and Shruti Aggarwal,"Performance Analysis of Association Rule Mining Algorithms",Volume 3, Issue 8, August 2013 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.
- [40] Supiyandi, Mochammad Iswan Perangin-angin, Andre Hasudungan Lubis, Ali Ikhwan, Mesran and Andysah Putera Utama Siahaan,"Association Rules Analysis on FP-Growth Method in Predicting Sales", DOI : 10.23883/IJRTER.2017.3453.DHCOA.
- [41] A. Singh and N. Jain,"Internet Surfing Prediction System using Association Rule Mining based on FP-Growth",International Journal of Scientific Research in Computer Science and Engineering, Vol.4, Issue.4, pp.1-6, August (2016).