

**MENERAPKAN DATA MINING DENGAN ALGORITMA FP-GROWTH PADA
ANALISIS POLA PEMBELIAN KONSUMEN PADA DONAT BOLONG
SEMARANG**

***THE APPLYING DATA MINING USING THE FP-GROWTH ALGORITHM IN THE
ANALYSIS OF CONSUMER PURCHASE PATTERNS ON DONUT BOLONG
SEMARANG***

Arya Sena Setyanegara¹, Eri Zuliarso²

^{1,2}Teknik Informatika Fakultas Teknologi Informasi dan Industri, Unisbank Semarang
arya.s.negara.3@gmail.com, eri299@edu.unisbank.ac.id

ABSTRACT

Competition in the business world is getting tougher day by day, so business people must make strategies to increase sales of their products. One thing that companies can do to carry out their business activities is utilize information technology. Technology that companies can use to design sales strategies is data mining. Donat Bolong Semarang is one of the fast food outlets in Semarang. This Bolong Donut presents a variety of flavours and toppings that are offered so that customers like it at low prices, which are very friendly for student pockets. This research will be carried out using the FP-Growth Algorithm, Kaggle tools, and the Python programming language. The results of data mining research that has been carried out using the FP-Growth Algorithm in purchasing transactions for Donuts Bolong Semarang can help the business strategy for Donuts Bolong Semarang promotion so that it can attract consumers.

Keywords: *Data Mining, FP-Growth Algorithm, Python, Kaggle*

ABSTRAK

Persaingan di dunia bisnis semakin hari semakin ketat sehingga para pelaku bisnis harus membuat strategi untuk meningkatkan penjualan produknya. Salah satu yang bisa dilakukan oleh perusahaan dalam menjalankan kegiatan bisnisnya adalah dengan cara memanfaatkan teknologi informasi. Teknologi yang bisa dimanfaatkan perusahaan untuk merancang strategi penjualan adalah dengan menerapkan penggunaan Data Mining. Donat Bolong Semarang adalah salah satu outlet makanan cepat saji yang berada di Semarang. Donat Bolong ini menyajikan berbagai macam rasa dan topping yang ditawarkan sehingga membuat para pelanggan menyukainya dengan harga murah yang sangat ramah untuk kantong pelajar ataupun mahasiswa. Pada penelitian ini akan dilakukan menggunakan Algoritma FP-Growth, tools Kaggle dan bahasa pemrograman python. Hasil dari penelitian data mining yang telah dilakukan menggunakan Algoritma FP-Growth pada transaksi pembelian Donat Bolong Semarang dapat membantu strategi bisnis pada Donat Bolong Semarang promosi sehingga dapat menarik konsumen.

Kata Kunci: Data Mining, Algoritma FP-Growth, Python, Kaggle

PENDAHULUAN

Persaingan di dunia bisnis semakin hari semakin ketat sehingga para pelaku bisnis harus membuat strategi untuk meningkatkan penjualan produknya. Salah satu yang bisa dilakukan oleh perusahaan dalam menjalankan kegiatan bisnisnya adalah dengan cara memanfaatkan teknologi informasi. Teknologi yang bisa dimanfaatkan perusahaan untuk merancang strategi penjualan adalah dengan menerapkan penggunaan Data Mining.

Data Mining adalah salah satu ilmu komputer yang sering digunakan untuk pengolahan data dan memberikan pola atau hubungan tertentu sehingga dihasilkan informasi yang bermanfaat bagi penggunaannya. Data Mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan mesin learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dalam database yang besar.

Pengelompokan Data Mining dibagi menjadi beberapa kelompok yaitu:

- a. Deskripsi
Deskripsi merupakan cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data yang dimiliki.
- b. Estimasi
Estimasi hampir sama dengan klasifikasi, kecuali variable target estimasi lebih ke arah numerik daripada ke arah kategori. Model yang dibangun menggunakan record lengkap yang menyediakan nilai variable target sebagai nilai prediksi.
- c. Prediksi
Prediksi menerka sebuah nilai yang belum diketahui dan juga memperkirakan nilai untuk masa mendatang.
- d. Klasifikasi
Dalam klasifikasi terdapat target variable kategori, misal penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu tinggi, sedang, dan rendah.
- e. Pengklasteran
Merupakan pengelompokan record, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.
- f. Asosiasi
Asosiasi bertugas menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

Donat Bolong Semarang adalah salah satu outlet makanan cepat saji yang berada di Semarang. Donat Bolong ini menyajikan berbagai macam rasa dan topping yang ditawarkan sehingga membuat para pelanggan menyukainya dengan harga murah yang sangat ramah untuk kantong pelajar ataupun mahasiswa.

Penelitian ini dilakukan untuk dapat menemukan pola pembelian donat dilakukan oleh pelanggan di Donat Bolong Semarang. Data yang digunakan pada penelitian ini adalah data transaksi dari penjualan donat yang dibeli oleh

pelanggan. Data penjualan akan dianalisis untuk mengetahui tingkat kecenderungan konsumen terhadap produk yang akan dibeli, sehingga akan diperoleh suatu pola beli konsumen terhadap produk tersebut.

Algoritma FP-Growth adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan data yang sering muncul (frequent itemset) dalam sebuah kumpulan data. Algoritma FP-Growth jika dibandingkan dengan algoritma lain dalam metode asosiasi seperti algoritma apriori, memiliki waktu kerja yang lebih cepat dalam menemukan frequent itemset.

Python adalah bahasa pemrograman yang digunakan oleh perusahaan besar yang mengembangkan berbagai aplikasi desktop, web, seluler, dan seluler. Python dikembangkan oleh Guido van Rossum di Belanda pada tahun 1990 dan dinamai menurut acara televisi favorit Guido Monty, Python's Flying Circus.

METODE

a. Pengumpulan Data

Tahapan ini merupakan persiapan awal dari penelitian ini. Dengan mendapatkan terlebih dahulu data pembelian dari Donat Bolong Semarang. Dengan melakukan observasi dan wawancara terhadap salah satu karyawan Donat Bolong Semarang.

1) Studi Observasi

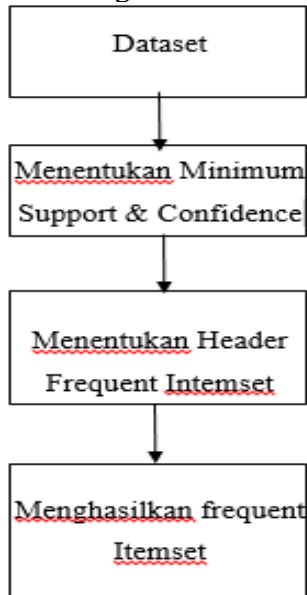
Studi observasi yaitu teknik atau pendekatan untuk mendapatkan data primer dengan cara mengamati langsung objek datanya. Pendekatan dilakukan dengan mengamati secara langsung kejadian pada sistem penjualan barang di Donat Bolong Semarang.

2) Studi Pustaka

Studi literatur yaitu studi yang dilakukan peneliti dalam mencari teori-teori yang bersangkutan dengan suatu penelitian yang dilakukan. Seperti teori-teori data mining, teori-metode-metode, ataupun teori-teori yang bersangkutan lainnya yang mana bisa berupa jurnal, majalah ataupun buku.

b. Penerapan Algoritma FP-Growth

Pada penelitian ini akan dibahas bagaimana cara mencari *frequent itemset* menggunakan algoritma FP-Growth.. Algoritma FP-Growth digunakan untuk menentukan *frequent itemset* (data yang paling sering muncul) dari sebuah dataset. Untuk menentukan *frequent itemset* pada data transaksi dapat digambarkan dalam diagram blok sebagai berikut:



Gambar 1. Frequent Itemset

c. Pemodelan Algoritma FP-Growth

Pada penelitian ini akan dilakukan pemodelan menggunakan Algoritma *Fp-Growth* untuk pengolahan data barang. Teknik data mining yang dipilih dalam penelitian ini adalah teknik asosiasi. *Association Rule Mining* adalah suatu prosedur untuk mencari hubungan antar *item* dalam suatu *dataset*. Dimulai dengan mencari *frequent itemset*, yaitu kombinasi yang paling sering terjadi dalam suatu *itemset* dan harus memenuhi *minsup*. [5] Pemodelan bertujuan untuk mencari aturan asosiasi, dimana aturan asosiasi selanjutnya dijadikan acuan untuk menentukan stok barang. Adapun langkah-langkah pembentukan model data mining dengan algoritma *Fp-Growth* adalah:

1. Menentukan data yang akan diproses.

Data yang di uji adalah data transaksi penjualan produk makanan di Toko Donat Bolong Semarang.

2. Menentukan minimum support dan minimum confidence.

d. Pengujian Algoritma FP-Growth

Pengujian dilakukan untuk mengetahui hasil perhitungan yang dianalisis dan untuk mengetahui apakah fungsi tersebut bekerja dengan baik atau tidak. Selanjutnya data tersebut diuji dengan menggunakan tools Kaggle untuk menghasilkan hasil penelitian

HASIL DAN PEMBAHASAN

Berikut merupakan Langkah – Langkah penelitian data mining yang telah dilakukan menggunakan algoritma *FP-Growth* :

1. Membuat Dataset

Langkah pertama dalam pembuatan dataset adalah menyiapkan data pembelian Donat Bolong Semarang. Kemudian data tersebut disimpan keformat *Comma Separated Values (CSV)*. Dataset adalah suatu database didalam memori (in-memory). Dataset adalah objek yang merepresentasikan data dan relasinya di memory. Strukturnya mirip dengan data yang ada di database.

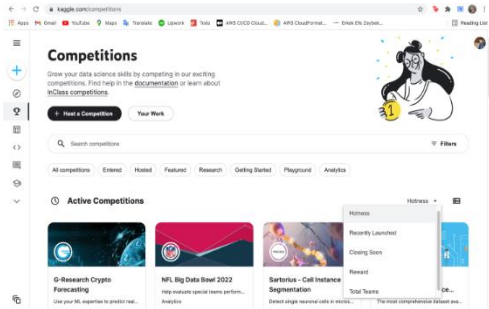
Tabel 1. Dataset

NO	VARIAN	KATEGORI	TRANSAKSI_ID
1	Classic	Regular	DNT-00001
2	Stroberi	Regular	DNT-00001
3	Oreo	Regular	DNT-00001
4	Taro	Regular	DNT-00001
5	Tiramisu	Regular	DNT-00002
6	Taro	Regular	DNT-00002
7	Martabak	Regular	DNT-00002
8	Keju	Regular	DNT-00002
9	Classic	Regular	DNT-00003
10	Frozen Pack	Paket	DNT-00003

2. Masuk Ke Website Kaggle

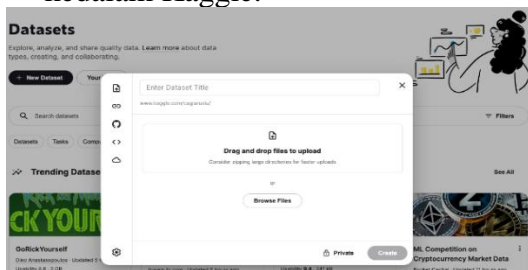
Kaggle adalah platform yang dapat diakses diberbagai macam *browser* seperti *Chrome, Opera, Mozilla Firefox* dll. *Kaggle* menggunakan bahasa pemrograman *R* atau *Python*. *Kaggle* sangat membantu para developer/data scientist/yang lainnya untuk mencari referensi kode-kode *machine learning*.

Berbagai lapisan yang ada dalam feature learning yang memiliki kegunaan untuk mentranslasikan sebuah input menjadi features.



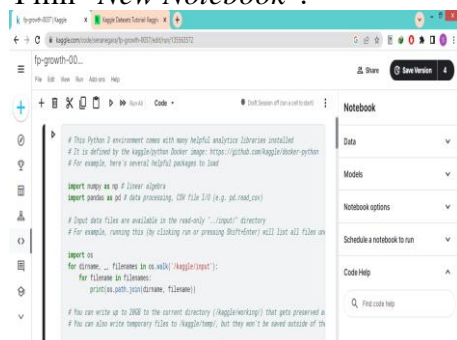
Gambar 2. Tampilan Kaggle

3. Mengupload Dataset
Beberapa langkah dalam mengupload dataset ke dalam Kaggle:
 1. Pilih menu “Dataset”.
 2. Kemudian klik “New Dataset”.
 3. Beri nama file dataset tersebut.
 4. Kemudian upload dataset anda kedalam Kaggle.



Gambar 3. Tampilan Upload Dataset

4. Tampilan Notebook Kaggle
Notebook adalah lembar kerja di Kaggle yang berfungsi sebagai lembar kerja bagi pengguna yang akan melakukan sebuah penelitian. Berikut ini cara membuka *notebook* di Kaggle :
 - a. Pilih “Create” yang ada di menu Kaggle.
 - b. Pilih “New Notebook”.



Gambar 4. Tampilan Notebook

5. Import Dataset
Masukan kode program :
data
pd.read_csv('/kaggle/input/dataset-skripsi-baru/Dataset_Skripsi_Baru.csv',header=None)
data

```
[8]:
```

	0	1	2	3
0	NO	VARIAN	KATEGORI	TRANSAKSI_ID
1	1	Classic	Regular	DNT-00001
2	2	Stroberi	Regular	DNT-00001
3	3	Oreo	Regular	DNT-00001
4	4	Taro	Regular	DNT-00001
...
737	737	Vanilla Almond	Regular	DNT-00214
738	738	Cokelat	Regular	DNT-00215
739	739	Keju	Regular	DNT-00215
740	740	Tiramisu	Regular	DNT-00215
741	741	Taro	Regular	DNT-00215

742 rows x 4 columns

Gambar 5. Tampilan Import Dataset

6. Preprocessing Data
Preprocessing data dilakukan menggunakan *TransactionEncoder()* yaitu mengubah tipe data format transaksi menjadi array dengan menggunakan kode dibawah ini :

#converting into required format of TransactionEncoder()

```
trans=[]
```

```
for i in range(0,741):
```

```
trans.append([str(data.values[i,j]) for j in range(0,3)])
```

```
trans=np.array(trans)
```

```
print(trans.shape)
```

Kemudian data array yang dihasilkan, kemudian diubah menjadi data tabular dengan menggunakan perintah dibawah ini :

```
## Using TransactionEncoder
```

```
t=TransactionEncoder()
```

```
data=t.fit_transform(trans)
```

```
data=pd.DataFrame(data,columns=t.columns_,dtype=int)
```

```
data.shape
```

```
data.head()
```

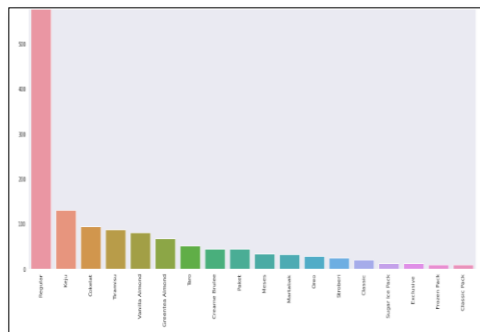
	Classic Pack	Cokelat	Creamme Brulee	DNT-0001	DNT-0002	DNT-0003	DNT-0004	DNT-0005	DNT-0006	Oreo Palet	Regular	Stroberi	Sugar Ice Pack	TRANSAKSI_ID	Taro	Tiramisu
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0
3	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0

Gambar 6. Tampilan Data Tabular

7. Menampilkan Jumlah Barang Yang Sering Dibeli

Dari semua transaksi akan ditampilkan 18 item yang memiliki frekuensi tertinggi menggunakan kode dibawah ini :

```
##Lets consider the top 18 items
purchased frequently
r=data.sum(axis=0).sort_values(asc
ending=False)[:18]
#altering the figsize
plt.figure(figsize=(20,10))
s=sns.barplot(x=r.index,y=r.values)
s.set_xticklabels(s.get_xticklabels()
, rotation=90)
```



Gambar 7. Tampilan Barang Yang Sering Dibeli

Pada Histogram diatas menunjukkan urutan 18 item yang sering dibeli oleh konsumen. Item yang sering dibeli adalah kategori “Regular”. Untuk nomor dua ada varian “Keju”, kemudian dinomor tiga ada varian “Cokelat”, diurutkan ke empat ada varian “Tiramisu”, diurutkan ke lima ada varian “Vanilla Almond”, diurutkan ke enam ada varian “Greentea Almond”, diurutkan ketujuh ada varian “Taro”, diurutkan kedelapan ada varian “Creamme Brulee”, diurutkan kesembilan ada

kategori “Paket”, diurutkan kesepuluh ada varian “Meses”, diurutkan kesebelas ada varian “Martabak”, diurutkan keduabelas ada varian “Oreo”, diurutkan ketigabelas ada varian “Stroberi”, diurutkan keempatbelas ada varian “Classic”, diurutkan kelimabelas ada varian “Sugar Ice Pack”, diurutkan keenambelas ada varian “Exclusive”, diurutkan ketujuhbelas ada varian “Frozen Pack” dan yang terakhir ada varian “Classic Pack”.

8. Peta Pohon

Untuk menampilkan 18 item yang paling sering dibeli dibuat menjadi peta dengan menggunakan kode dibawah ini :

```
#treemap of top 18 frequent items
plt.figure(figsize=(15,15))
squarify.plot(sizes=r.values,
label=r.index, alpha=.7,color=colors)
plt.title("Tree map of top 18 items")
plt.axis('off')
```



Gambar 8. Tampilan Peta Pohon

9. Menampilkan Frequent Itemset

Untuk mencari frequent itemset-nya dengan menggunakan kode dibawah ini :

```
#let us return items and itemsets
with atleast 5% support:
freq_items=fpgrowth(data,min_sup
port=0.005,use_colnames=True)
freq_items
```

	support	itemsets
0	0.939271	(Regular)
1	0.026991	(Classic)
2	0.005398	(DNT-00001)
3	0.033738	(Stroberi)
4	0.037787	(Creo)
...
185	0.006748	(DNT-00209, Regular)
186	0.005398	(DNT-00211, Regular)
187	0.005398	(Regular, DNT-00212)
188	0.006748	(DNT-00213, Regular)
189	0.006748	(DNT-00214, Regular)

190 rows x 2 columns

Gambar 9. Frequent Itemset

10. Menampilkan Asosiasi

Untuk menampilkan asosiasi dari transaksi pembelian Donat Bolong Semarang maka digunakanlah kode dibawah ini :

```
res=association_rules(freq_items,metric="support",min_threshold=0.0001)
res
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(Regular)	(Classic)	0.939271	0.026991	0.026991	0.028736	1.064553	0.001439	1.001797
1	(Classic)	(Regular)	0.026991	0.939271	0.026991	1.000000	1.064553	0.001439	inf
2	(Paket)	(Classic Pack)	0.059379	0.012146	0.012146	0.204545	16.840939	0.011425	1.241674
3	(Classic Pack)	(Paket)	0.012146	0.059379	0.012146	1.000000	16.840939	0.011425	inf
4	(Cokelat)	(Regular)	0.128205	0.939271	0.128205	1.000000	1.064553	0.007786	inf
...
177	(Regular)	(Taro)	0.939271	0.066826	0.066826	0.073276	1.064553	0.004180	1.004802
178	(Regular)	(Tiramisu)	0.939271	0.117409	0.117409	0.125000	1.064553	0.007130	1.008676
179	(Tiramisu)	(Regular)	0.117409	0.939271	0.117409	1.000000	1.064553	0.007130	inf
180	(Regular)	(Vanilla Almond)	0.939271	0.108912	0.108912	0.116379	1.064553	0.006636	1.007999
181	(Vanilla Almond)	(Regular)	0.108912	0.939271	0.108912	1.000000	1.064553	0.006636	inf

182 rows x 9 columns

Gambar 10. Hasil Asosiasi

Berikut merupakan hasil dari rincian frequent itemset dengan menggunakan minimum threshold 0.0001 :

1. Jika pelanggan membeli Regular maka pelanggan akan membeli Classic dengan nilai support 0.02 dan confidence 0.02.
2. Jika pelanggan membeli Classic maka pelanggan akan membeli Regular dengan nilai support 0.02 dan confidence 1.
3. Jika pelanggan membeli Paket maka pelanggan akan membeli Classic Pack

dengan nilai support 0.01 dan confidence 0.01.

4. Jika pelanggan membeli Regular maka pelanggan akan membeli Cokelat dengan nilai support 0.01 dan confidence 1.
5. Jika pelanggan membeli Regular maka pelanggan akan membeli Taro dengan nilai support 0.06 dan confidence 0.07.
6. Jika pelanggan membeli Tiramisu maka pelanggan akan membeli Regular dengan nilai support 0.01 dan confidence 1.
7. Jika pelanggan membeli Regular maka pelanggan akan membeli Tiramisu dengan nilai support 0.01 dan confidence 0.01.
8. Jika pelanggan membeli Vanilla Almond maka pelanggan akan membeli Regular dengan nilai support 0.01 dan confidence 1.

Jika pelanggan membeli Regular maka pelanggan akan membeli Vanilla Almond dengan nilai support 0.01 dan confidence 0.01

SIMPULAN

Dari penelitian data mining yang telah dilakukan menggunakan algoritma FP-Growth pada transaksi pembelian Donat Bolong Semarang, maka dapat ditarik kesimpulan sebagai berikut :

1. Dari kombinasi barang tersebut, maka dapat membantu Donat Bolong Semarang promosi sehingga dapat menarik konsumen.
2. Dari penelitian data mining yang telah dilakukan memanfaatkan tools Kaggle, menggunakan algoritma FP-Growth dan menggunakan bahasa pemrograman python maka dapat ditemukan 18 barang yang sering dibeli pada Donat Bolong Semarang.
3. Dari hasil penelitian yang diperoleh dengan menerapkan algoritma FP-Growth maka peneliti dapat menemukan barang yang sering dibeli bersamaan oleh konsumen.

DAFTAR PUSTAKA

- N. R. Ardani and N. Fitriana, "Sistem Rekomendasi Pemesanan Sparepart Dengan Algoritma Fp-Growth (Studi Kasus Pt. Rosalia Surakarta)," *Semin. Nas. Teknol. Inf. dan Multimed.* 2016, pp. 1–6, 2016, [Online]. Available: Rekomendasi, sparepart, Assosiation Rule, FP-Growth.%0A1.
- A. Maulana and A. A. Fajrin, "Penerapan Data Mining Untuk Analisis Pola Pembelian Konsumen Dengan Algoritma Fp-Growth Pada Data Transaksi Penjualan Spare Part Motor," *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 1, p. 27, 2018, doi: 10.20527/klik.v5i1.100.
- Vulandari Tri Retno, "Data Mining Teori dan Aplikasi Rapidminer," p. viii+124, 2017, [Online]. Available: www.infogavamedia@yahoo.com
- Muhammad Romzi and B. Kurniawan, "Pembelajaran Pemrograman Python Dengan Pendekatan Logika Algoritma," *JTIM J. Tek. Inform. Mahakarya*, vol. 3, no. 2, pp. 37–44, 2020.
- F. Fatihatul, A. Setiawan, and R. Rosadi, "Asosiasi Data Mining Menggunakan Algoritma FP-Growth Untuk Market Basket Analysis," *Jatinangor Univ. Padjadjaran*, pp. 1–8, 2016.
- S. Suhartini, L. Kerta Wijaya, and N. Arini Pratiwi, "Penerapan Algoritma K-Means Untuk Pendataan Obat Berdasarkan Laporan Bulanan Pada Dinas Kesehatan Kabupaten Lombok Timur," *Infotek J. Inform. dan Teknol.*, vol. 3, no. 2, pp. 147–156, 2020, doi: 10.29408/jit.v3i2.2315.