

PENGELOMPOKAN DATA PENJUALAN SEMBAKO BERDASARKAN PERILAKU PEMBELI MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING

Muhammad Husen Adlun, Rini Astuti, Fadhil M Basysyar

Teknik Informatika, STMIK IKMI Cirebon
Jl. Perjuangan No. 10 B, Cirebon, Jawa Barat
husenadlun@gmail.com

ABSTRAK

Sembako adalah komoditas yang sangat penting bagi masyarakat. Oleh karena itu, distribusi dan penjualan sembako memiliki peran yang signifikan dalam perekonomian. Supermarket dan toko kelontong berupaya menyediakan produk yang paling diminati konsumen, sambil menyesuaikan strategi penjualan mereka sesuai dengan permintaan dan perilaku pembeli. Tujuan penelitian ini adalah untuk menganalisis perilaku pembeli dalam data penjualan sembako dan mengelompokkan data tersebut berdasarkan pola pembelian yang sebanding dengan algoritma pengelompokan K-Means. Diharapkan bahwa pemahaman yang lebih baik tentang demografi pembeli dan pola belanja mereka dapat dihasilkan dari penggunaan kelompok ini dalam strategi pemasaran. Data yang digunakan dalam penelitian ini berasal dari penjualan sembako dari toko selama satu tahun. Informasi tentang jenis produk, jumlah, waktu, dan identitas pembeli termasuk dalam data ini. Hasil dari Clustering ini menunjukkan bahwa hasil perhitungan dan uji coba dari penelitian yang sudah dilakukan, terdapat dua cluster yang telah ditentukan, untuk cluster_0 yang memiliki 1237 item dikategorikan laku, dan untuk cluster_1 memiliki 51 item dikategorikan kurang laku. Dengan nilai Davies Bouldin Index (DBI) 0.234.

Kata Kunci: K-Means, Clustering, Davies Bouldin Index (DBI)

1. PENDAHULUAN

Pelanggan adalah salah satu faktor yang sangat penting dalam kegiatan bisnis. Bagaimana sebuah bisnis atau industri memperhatikan pelanggannya, baik dalam hal pemenuhan produk, layanan penjualan, dan promosi, serta elemen lainnya, sangat memengaruhi keberlangsungan dan keberhasilan sebuah bisnis atau industri. Perusahaan tidak dapat menjual dan mengirimkan barang yang dibuat tanpa pelanggan. Artinya, jika produk tersebut tidak terjual, perusahaan tidak akan mendapatkan penghasilan, yang pada akhirnya akan mempengaruhi keberlangsungan bisnis.[1]

Dengan kemajuan dalam informasi dan analisis data, ada kesempatan baru untuk memahami perilaku konsumen dan mengembangkan strategi pemasaran yang lebih efektif. Dalam penjualan sembako, memahami perilaku pelanggan adalah penting untuk meningkatkan penjualan dan efisiensi operasi. Algoritma K-Means Clustering adalah salah satu cara untuk menemukan pola dan tren dalam data penjualan.

Sektor penjualan sembako adalah kebutuhan pokok bagi masyarakat, penjualan sembako memainkan peran penting dalam ekonomi Indonesia. Kajian ini mencoba memberikan wawasan baru tentang perilaku pembeli dalam hal penjualan sembako melalui analisis data penjualan sembako untuk menemukan pola pembelian yang dapat digunakan untuk membuat strategi bisnis yang lebih efektif.

Tujuan utama dari penelitian ini adalah untuk mengelompokkan data penjualan sembako dengan mempertimbangkan perilaku pembeli. Penelitian ini bertujuan untuk menemukan kelompok pembeli dengan pola dan karakteristik pembelian yang sebanding dengan algoritma Clustering K-Means.

Tujuan ini mendorong analisis lebih lanjut tentang strategi pemasaran yang lebih khusus, penyusunan produk, dan pelayanan pelanggan.

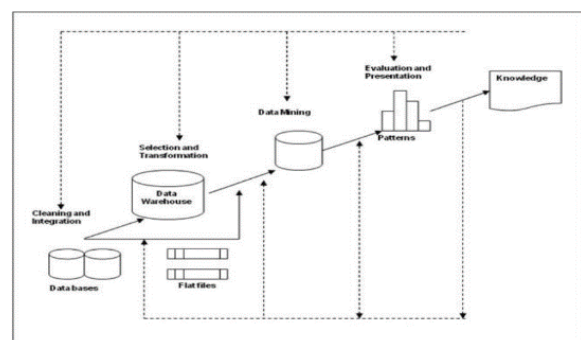
2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah istilah yang digunakan untuk menggambarkan proses menemukan pengetahuan tersembunyi dalam database. Data mining merupakan proses semi-otomatis yang menggabungkan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak serta mengidentifikasi informasi yang memiliki potensi dan nilai guna dari database berukuran besar.[2]

2.2. Knowledge Discovery in Database

Penelitian ini menggunakan metode Knowledge Discovery in Database, yang merupakan proses untuk mendapatkan informasi dari data tersembunyi dengan mengubahnya menjadi pola-pola karakteristik yang dapat diidentifikasi dari data tersebut.[3]



Gambar 1. Tahapan Proses KDD

1. Data Cleaning
2. Data Integration
3. Data Selection
4. Data Transformation
5. Data Mining
6. Evaluasi
7. Knowledge Presentation

2.3. K-Means

K-Means adalah salah satu algoritma clustering yang paling populer dan sederhana. Algoritma ini bekerja dengan mengelompokkan data ke dalam sejumlah klaster yang telah ditentukan sebelumnya. Proses K-Means dimulai dengan memilih jumlah klaster yang diinginkan (k) dan menentukan titik awal yang akan digunakan sebagai centroid untuk setiap klaster. Selanjutnya, setiap data diklasifikasikan ke dalam klaster yang centroid-nya paling dekat dengan data tersebut. Setelah semua data diklasifikasikan, centroid dari setiap klaster diperbarui berdasarkan rata-rata dari semua data yang termasuk dalam klaster tersebut. Proses ini diulang beberapa kali sampai tidak ada perubahan dalam pengelompokan data.[4]

2.4. Clustering

Clustering adalah suatu proses di mana data dikelompokkan menjadi beberapa kelompok dari sebuah kumpulan data yang tidak memiliki kelas yang diketahui sebelumnya, dan menentukan keanggotaan data dalam kelompok-kelompok tersebut. Potensi penggunaan clustering terletak pada kemampuannya untuk mengidentifikasi struktur dalam data, yang dapat diterapkan dalam berbagai aplikasi seperti pengenalan pola, pengolahan gambar, dan klasifikasi.[5]

2.5. Rapidminer

RapidMiner adalah perangkat lunak yang tersedia untuk umum dan bersifat open source. RapidMiner digunakan sebagai alat untuk menganalisis data secara menyeluruh. Dalam RapidMiner, berbagai teknik seperti deskriptif dan prediksi digunakan. Pengoperasian RapidMiner didasarkan pada bahasa pemrograman Java.[6]

2.6. Penelitian Terdahulu

Pada penelitian yang dilakukan oleh Bayu Rimbi Asmoro, Agus Wibowo, dan Ari Fiqri Aryadi ini berjudul “Penggunaan Algoritma K-Means Untuk Menganalisa Penjualan Di Bigmart” Dalam penelitian ini, algoritma yang digunakan K-Means untuk menganalisis penjualan di BigMart dibahas. Dalam penelitian ini, produk dikelompokkan menjadi tiga kelompok berdasarkan tingkat kelarisan. Software Rapidminer dan algoritma K-Means digunakan untuk menganalisis data penjualan BigMart. Penelitian mencakup proses preprocessing data menggunakan transformasi Z, penggunaan clustering K-Means untuk mengelompokkan data penjualan, dan penggunaan software Rapidminer. Hasil analisis menunjukkan

bahwa ada empat kategori produk yang paling laku, enam kategori produk yang laku, dan enam kategori produk yang tidak laku. Hasil evaluasi cluster menunjukkan bahwa cluster tersebut memiliki kualitas yang cukup, dengan nilai Davies Bouldin Index (DBI) sebesar 0,497.[7]

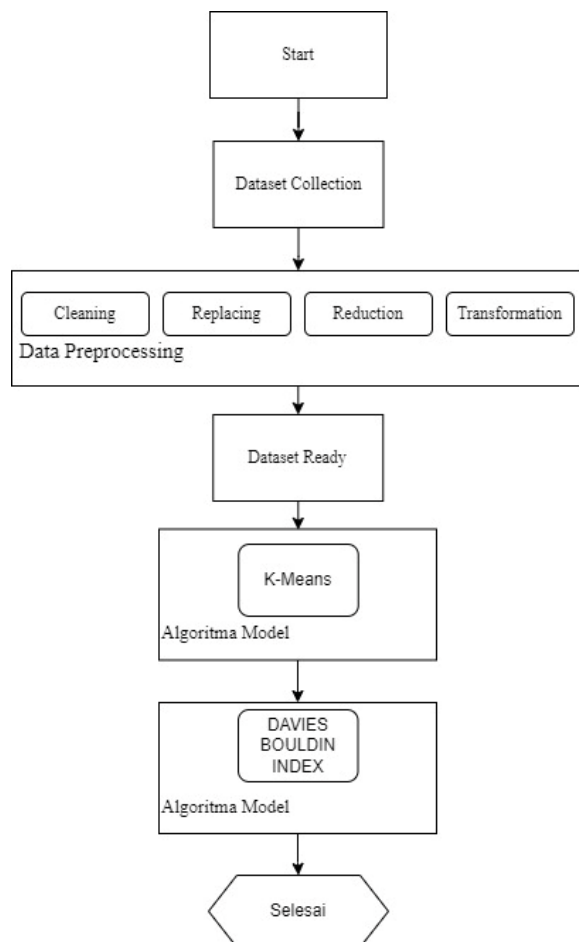
Pada penelitian yang dilakukan oleh Kasini, dan Nani Hidayati ini berjudul “Penerapan Data Mining Untuk Clustering Pada Toko Laura Grosir Dan Eceran Menggunakan Algoritma K-Means” penelitian ini membahas menggunakan metode clustering K-Means pada toko grosir dan eceran Laura. Tujuan dari penelitian ini adalah untuk mengelompokkan produk yang sangat laris, laris, dan kurang laris berdasarkan data penjualan selama tiga bulan terakhir. Data dari tujuh puluh lima produk dari toko tersebut digunakan dalam penelitian ini. Data produk dikelompokkan ke dalam tiga kelompok menggunakan metode pengelompokan K-Means berdasarkan karakteristik atau fitur yang sebanding. Proses ini dilakukan secara manual dan menggunakan software RapidMiner. Hasil penelitian menunjukkan bahwa terdapat tiga kelompok barang: barang yang sangat laris (3 barang), barang yang laris (9 barang), dan barang yang kurang laris (63 barang).[8]

Pada penelitian yang dilakukan oleh Putri Dwi Lestari, dan Mulyawan ini berjudul “Data Mining Pada Penjualan Air Bersih Di Spam Akidah Menggunakan Algoritma K-Means Clustering Menggunakan Rapidminer” Penelitian ini membahas penggunaan data mining untuk penjualan air bersih di Desa Kecomberan SPAM Akidah menggunakan algoritma clustering K-Means. Dengan menggunakan data kubikasi dari Januari hingga Desember 2022 milik 241 pelanggan, tujuan penelitian adalah untuk mengelompokkan pelanggan ke dalam dua kategori: pengguna hemat dan boros. Hasilnya menunjukkan bahwa 190 pelanggan termasuk dalam kategori irit di Cluster 0 dan 51 pelanggan termasuk dalam kategori boros di Cluster 1.[9]

Pada penelitian yang dilakukan oleh Yuli Hartati, Sarjon Defit, dan Gunadi Widi Nurcahyo ini berjudul “Klasterisasi Bibit Terbaik Menggunakan Algoritma K-Means dalam Meningkatkan Penjualan” penelitian ini membahas bagaimana algoritma K-Means digunakan untuk menggabungkan data penjualan bibit terbaik di UD. Tiara Brothers untuk meningkatkan penjualan. Studi ini menganalisis data penjualan bibit dari Januari hingga April 2019 dan menemukan 2 cluster yaitu bibit laris dan tidak laris, dengan 4 bibit unggul yang dapat meningkatkan penjualan. Data yang diuji terdiri dari 42 data bibit dari 2 cluster adalah 4 data bibit laris di cluster satu (C1), dan 38 data bibit tidak laris di cluster dua (C2).[10]

3. METODE PENELITIAN

Berikut ini klustering penjualan sembako pada penelitian ini menggunakan tahapan yang sistematis dan terarah, pada gambar berikut:



Gambar 2. Tahapan Metode Penelitian

Berikut penjelasan masing-masing tahapan dari metode penelitian, yaitu:

3.1. Pengumpulan Data

Dataset yang digunakan penelitian ini adalah dataset penjualan sembako yang diambil dari Kaggle dan berjumlah 1288 sampel. Pengumpulan data yang dirancang untuk memenuhi kebutuhan penelitian. Diharap bahwa data ini akan memungkinkan pengumpulan data yang relevan.

A	B	C	D	E	F
1	Tanggal	Nama pembeli	Nama Barang	Kuantum	Nominal
2	26/03/2020	YAYASAN KESEJAHTERAAN PEGAWAI (YKP) BNI	BERAS	2	46000
3	12/08/2020	SAHABAT ANDA	BERAS	3	69000
4	18/07/2020	YAYASAN KESEJAHTERAAN PEGAWAI (YKP) BNI	BERAS	4	92000
5	28/04/2020	TOKO DARUM KOPERASI DISPERINDAG	BERAS	4	92000
6	28/04/2020	YAYASAN KESEJAHTERAAN PEGAWAI (YKP) BNI	BERAS	4	92000
7	13/02/2020	TOKO RIZKI SAKYAH	TEPUNG	20	166000
8	24/08/2020	TOKO SIDAQ	TEPUNG	20	166520
9	8/02/2020	TOKO MOMONG ALPINA	TEPUNG	20	166520
10	10/12/2020	TOKO SIDAQ	TEPUNG	20	166520
11	10/09/2020	TOKO SIDAQ	TEPUNG	20	166520
12	11/01/2021	TOKO SIDAQ	TEPUNG	20	166520
13	4/06/2020	TOKO RIZKI SAKYAH	TEPUNG	40	320000
14	20/01/2020	TOKO AGIL	TEPUNG	40	320000
15	03/02/2020	TOKO DARUM KOPERASI DISPERINDAG	TEPUNG	40	320000
16	17/02/2020	TOKO SILVIANA	TEPUNG	40	320000
17	06/03/2020	TOKO MAJU TERUS	TEPUNG	40	320000
18	22/03/2020	Warung Nak Zihan	TEPUNG	40	320000
19	22/10/2020	TOKO ERA SAPUTRA JAYA	TEPUNG	40	330400
20	9/12/2020	TOKO YUSMAN DABU	TEPUNG	40	330400
21	21/12/2020	TOKO BU YULI	TEPUNG	40	330400

Gambar 3. Dataset Penjualan Sembako

Berikut penjelasan atribut pada gambar 3:

- Kuantum: Bisa kita sebut jumlah pembelian.
- Nominal: Bisa kita sebut jumlah transaksi pembelian.

1. Preprocessing Data

Data dalam penelitian dikumpulkan dan diidentifikasi, lalu dilakukan preprocessing, yaitu pembersihan data (cleaning) dan replacing (missing values). Setelah tahap preprocessing, data telah siap untuk digunakan untuk tahap berikutnya.

2. Penerapan Algoritma

Penelitian ini menggunakan algoritma K-Means untuk mendapatkan hasil dari kumpulan data yang akan dibagi menjadi 3 cluster.

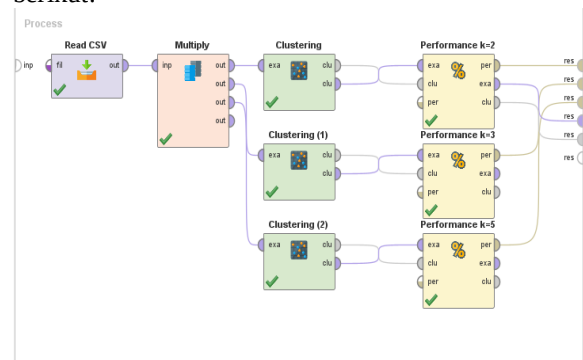
3. Evaluasi

Proses evaluasi akurasi algoritma pada penelitian dilakukan dengan menggunakan model Davies Bouldin Index (DBI) untuk melihat bagaimana hasil dari model atau metode yang digunakan dibandingkan dengan hasil sebelumnya.

4. HASIL DAN PEMBAHASAN

4.1. Penerapan Clustering pada Rapidminer

Desain model Algoritma K-Means yang menggunakan aplikasi Rapidminer adalah sebagai berikut:



Gambar 4. Tampilan Pemodelan K-Means Clustering

Ada 3 proses yang dilakukan, yaitu:

a. Read CSV

Tahapan ini dilakukan untuk menginput dataset berupa file berekstensi CSV dataset penjualan sembako.

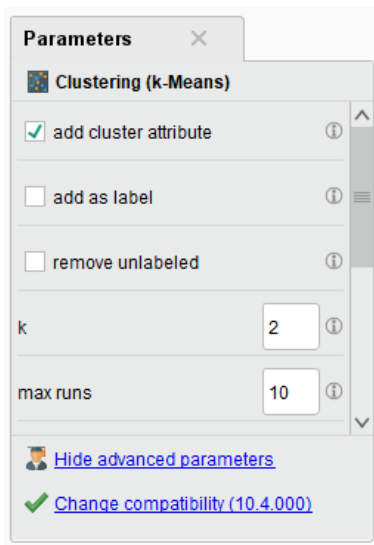
b. Clustering

Tahapan ini melakukan operasi cluster sebagai algoritma yang akan dilakukan pada penelitian ini.

c. Performance

Pada tahap ini melakukan operasi pencarian nilai Davies Bouldin Index (DBI).

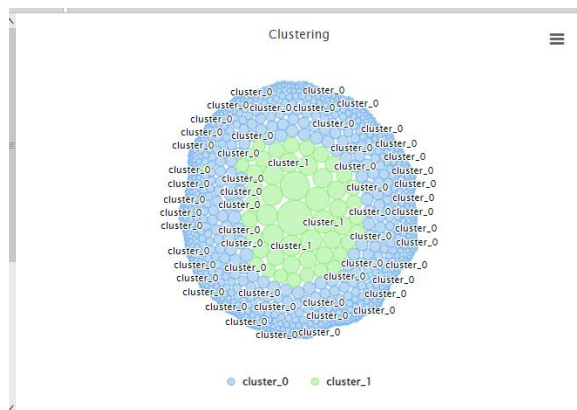
Parameter yang digunakan untuk proses cluster adalah sebagai berikut:



Gambar 5. Parameter K-Means

Gambar Parameter K-Means yang digunakan. Pada gambar 5, nilai K merupakan nilai yang digunakan untuk menentukan jumlah cluster yang akan dibentuk.

Setelah dilakukan dan melewati tiga tahapan di atas, rapidminer menghasilkan beberapa output dan hasil pengujian, yang terlihat sebagai berikut:



Gambar 6. Visualisasi Grafik Cluster Dengan Packed Bubble

Hasil pengelompokan data ditunjukkan dalam gambar 6 visualisasi data dalam bentuk grafik packed bubble 2 warna. Warna biru menunjukkan cluster_0 yang memiliki 1237 item, dan warna hijau menunjukkan cluster_1 yang memiliki 51 item.

Adapun titik pusat setiap cluster pada gambar 7, Dari hasil penerapan algoritma K-Means pada data penjualan sembako dengan menggunakan rapidminer diperoleh nilai centroid yang ditunjukkan di gambar 7 Centroid Table.

Attribute	cluster_0	cluster_1
Kuantum	1127.407	12545.490
Nominal	12855376.582	138947127.451

Gambar 7. Centroid Table

4.2. Hasil Penelitian

Berikut adalah hasil dari perhitungan dari tahapan-tahapan metode, yaitu:

1. Pengulangan yang dilakukan k=2, k=3, dan k=5 berikut hasilnya:

Tabel 1. hasil nilai DBI masing-masing K.

NO	Kluster	Nilai DBI
1	K=2	0.234
2	K=3	0.248
3	K=5	0.240

Berdasarkan nilai Davies Bouldin Index (DBI) dalam penelitian ini mendapatkan nilai K=2 adalah 0.234, nilai K=3 adalah 0.248, dan nilai K=5 adalah 0.240. maka nilai terkecil dari cluster tersebut adalah K=2 dengan nilai 0.234.

2. Berdasarkan hasil perhitungan dan uji coba dari tahapan-tahapan metode yang sudah dilakukan, terdapat 2 cluster yang telah ditentukan, untuk cluster_0 yang memiliki 1237 item dikategorikan laku, dan untuk cluster_1 memiliki 51 item dikategorikan kurang laku.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil menemukan pola dalam pembelian sembako dengan menganalisis perilaku konsumen menggunakan algoritma K-Means Clustering. Hasil dari pengelompokan data penjualan sembako yang diambil dari platform Kaggle mengungkapkan bahwa perilaku pembeli dapat dikelompokkan ke dalam 2 kelompok besar, di mana setiap cluster memiliki karakteristik dan pola pembelian yang berbeda. Berdasarkan hasil dari penelitian yang dilakukan, dapat disimpulkan sebagai berikut: Terdapat 2 cluster dengan sebaran data sebanyak 1237 pembeli dikategorikan laku, dan untuk kategori kurang laku sebanyak 51 pembeli. Nilai Davies Bouldin Index yang dihasilkan dari proses cluster distance performance cluster_0 sebesar 0.234.

Untuk penelitian berikutnya, direkomendasikan untuk melakukan langkah-langkah sebagai berikut: membandingkan algoritma k-Means dengan pendekatan hybrid k-Means, memperluas jumlah data yang diproses, dan menambah kriteria penilaian, seperti nilai piutang pelangan.

DAFTAR PUSTAKA

- [1] W. Sudrajat, I. Cholid, M. Rachmadi, and V. Amalia, "Penerapan Algoritma K-Means Dalam Segmentasi Pelanggan Pada Toko Sembako Menggunakan Rapidminer Application of the K-Means Algorithm in Customer Segmentation at Grocery Stores Using Rapidminer."
- [2] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "PENERAPAN DATA MINING METODE K-MEANS CLUSTERING UNTUK ANALISA PENJUALAN PADA TOKO YANA SPORT," 2022.
- [3] A. Muhidin and S. Kevin Alfandara, "ANALISIS DATA MINING MENGGUNAKAN ALGORITMA K-MEANS UNTUK CLUSTERING PENJUALAN STUDI KASUS DAPUR BU IPUNG," vol. 13, 2022.
- [4] M. R. Nahjan, N. Heryana, and A. Voutama, "IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL," 2023.
- [5] Y. Dharma Putra, M. Sudarma, and I. B. A. Swamardika, "Clustering History Data Penjualan Menggunakan Algoritma K-Means," *Majalah Ilmiah Teknologi Elektro*, vol. 20, no. 2, p. 195, Dec. 2021, doi: 10.24843/mite.2021.v20i02.p03.
- [6] Y. Ratna Sari, A. Sudewa, D. Ayu Lestari, and T. Ika Jaya, "PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING DATA KEMISKINAN PROVINSI BANTEN MENGGUNAKAN RAPIDMINER," 2020.
- [7] B. R. Asmoro, A. Wibowo, and A. F. Aryadi, "PENGGUNAAN ALGORITMA K-MEANS UNTUK MENGANALISA PENJUALAN DI BIGMART," 2022.
- [8] Kasini and N. Hidayati, "Penerapan Data Mining Untuk Clustering Pada Toko Laura Grosir Dan Eceran Menggunakan Algoritma."
- [9] P. D. Lestari, "DATAMINING PADA PENJUALAN AIR BERSIH DI SPAM AKIDAH MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING MENGGUNAKAN RAPIDMINER," 2023.
- [10] Y. Hartati, S. Defit, and G. W. Nurcahyo, "Klasterisasi Bibit Terbaik Menggunakan Algoritma K-Means dalam Meningkatkan Penjualan," *Jurnal Informatika Ekonomi Bisnis*, Mar. 2021, doi: 10.37034/infeb.v3i1.56.